Word count: 10403 words (main body of text including references).

Type of article: Original Article.

Title: Phylogeographic structure of the pygmy shrew: revisiting the roles of southern and northern refugia in Europe.

Authors: RODRIGO VEGA1*, ALLAN D. MCDEVITT2, JOANNA STOJAK3, ALINA MISHTA4, JAN M. WÓJCIK3, BORIS KRYŠTUFEK5 and JEREMY B. SEARLE6

Affiliations: 1Ecology Research Group, Section of Natural and Applied Sciences, School of Human and Life Sciences, Canterbury Christ Church University, Becket Bf23, North Holmes Road, Canterbury, Kent, CT1 1QU, UK.
2Ecosystems and Environment Research Centre, School of Science, Engineering and Environment, University of Salford, Salford, M5 4WT, UK.
3Mammal Research Institute, Polish Academy of Sciences, 17-230 Białowieża, Poland.
4Schmalhausen Institute of Zoology, National Academy of Sciences of Ukraine, 01601 Kiev, Ukraine.
5Slovenian Museum of Natural History, Prešernova 20, SI-1000 Ljubljana, Slovenia.
6Department of Ecology and Evolutionary Biology, Cornell University, Ithaca, New York, 14853-2701, USA.

*Corresponding author. E-mail: rodrigo.vega@canterbury.ac.uk
Running title:
Phylogeography of the pygmy shrew.
ABSTRACT

Southern and northern glacial refugia are considered paradigms that explain the complex phylogeographic patterns and processes of European biota. Here, we provide a revisited statistical phylogeographic analysis of the pygmy shrew Sorex minutus Linnaeus, 1766 (Eulipotyphla, Soricidae) examining the genetic diversity, genetic differentiation and demographic history in the Mediterranean peninsulas and in Western and Central Europe. The results showed support for genetically distinct and diverse phylogeographic groups consistent with southern and northern glacial refugia, as expected from previous studies, but also identified geographical barriers concordant with glaciated mountain ranges during the Last Glacial Maximum (LGM), early diversification events dated between the Upper Pleistocene and Lower Holocene for the main phylogeographic groups, and recent (post-LGM) patterns of demographic expansions. This study is the most comprehensive investigation of this species to date, and the results have implications for the conservation of intraspecific diversity and the preservation of the evolutionary potential of S. minutus.

INTRODUCTION

During the Quaternary glaciations, species in Europe were restricted to glacial refugia at glacial maxima (Bilton et al., 1998; Taberlet et al., 1998; Hewitt, 2000; Stewart & Lister, 2001; Pazonyi, 2004; Sommer & Nadachowski, 2006). As glaciers retreated, a broad range of recolonisation patterns emerged, as evidenced by palaeontological, biogeographic and phylogeographic studies on various taxa, resulting in the complex contemporary patterns of endemism, species richness and biodiversity hotspots observed across Europe. While population contraction and lineage diversification within southern glacial refugia in the Mediterranean peninsulas during the Last Glacial Maximum [LGM; 19-26.5 thousand years ago (KYA) (Clark et al., 2009)], and subsequent northward postglacial recolonisation of Europe have been accepted and recognised since the 1990s (Bilton et al., 1998; Taberlet et al., 1998; Hewitt 2000), the concept of northern glacial refugia also became a paradigm to explain the complex phylogeographic patterns and processes of European biota (Stewart & Lister, 2001; Pazonyi 2004; Sommer & Nadachowski, 2006). Fossil records and phylogenetic analyses revealed that many species of flora and fauna could have survived during the LGM in the Carpathian Basin (Stewart & Lister, 2001; Pazonyi, 2004; Sommer & Nadachowski, 2006; Stojak et al., 2015) and in the Dordogne region (Steward et al., 2010), and glacial refugia could also be located in Crimea (Marková, 2011) or in the Russian Plain (Banaszek et al., 2012). Nowadays, locations of southern and northern glacial refugia during the LGM are hotspots of genetic diversity (Petit et al., 2003; Stojak et al., 2016).

The Eurasian pygmy shrew Sorex minutus Linnaeus, 1766 (Eulipotyphla, Soricidae) (Hutterer, 1990) has been used as a phylogeographic model species for understanding the effects of the glaciations in Europe and the colonisation history during the Pleistocene and postglacial times (Bilton et al., 1998: McDevitt et al., 2010; Vega et al., 2010a, b). However, little is still known about the phylogeographic structure, genetic diversity and structure, and demographic history of this small mammal within these regions due to the limited number of samples from Mediterranean peninsulas. An expanded phylogeographic study of the pygmy shrew is therefore important for the understanding and further development of biogeographic
models of glacial refugia and postglacial recolonization, for depicting areas with high intraspecific genetic diversity, for establishing conservation measures of rear-edge populations, and for the preservation of the evolutionary potential of species, particularly in the face of climate and anthropogenic change (Deffontaine et al., 2005; Provan & Bennett, 2008; Stojak et al., 2019; Stojak & Tarnowska, 2019).

In this study, we explored the evolutionary history and phylogeographic structure of *Sorex minutus* using a statistical phylogeography approach (Knowles & Maddison, 2002; Knowles, 2009). Here, we emphasise the genetic diversity and structure within and among refugia, the inference of geographical barriers and the demographic history of *S. minutus*, which are aspects that have not been studied in detail previously. Specifically, we asked the following questions: 1) What are the geographical distribution and genetic diversity patterns of the genealogical lineages of *S. minutus*? 2) Is there significant population genetic structure across the geographic range of *S. minutus*? 3) What is the historical demography of *S. minutus* in Europe? Our results showed support for distinct and genetically diverse lineages, geographical barriers concordant with glaciated mountain ranges during the LGM, and recent (post-LGM) population expansions with contemporary contact areas. The results presented here have implications for the long-term conservation of intraspecific diversity and the preservation of the evolutionary potential of *S. minutus* in the face of modern climate change.

**MATERIALS AND METHODS**

**Study species**

*Sorex minutus* is common over most of its distribution but is rarely dominant and it occurs in a wide range of terrestrial habitats with adequate ground cover and in relatively damp areas, including swamps, grasslands, heaths, sand dunes, woodland edge, rocky areas, shrubland and montane forests (Hutterer, 1990, 2016; Churchfield, 1990; Churchfield & Searle, 2008). It is found from southern and western Europe to much of central and northern Europe, Ireland and the British Isles, and Siberia to Lake Baikal in the east (Hutterer, 1990, 2016). It
is found from sea level up to 2260 m (in the Alps), but its distribution becomes patchy and
limited to higher altitudes in southern Europe where it occurs with some degree of
geographical isolation and differentiation, while in central and northern parts of Europe and
in Siberia it is more abundant and populations are more connected and widespread
(Hutterer, 1990, 2016).

**Samples and molecular methods**

A total of 671 mitochondrially encoded cytochrome b (*MT-CYB*) DNA sequences of *S.
*minutus* from Europe and Siberia were used for this study (Fig. 1B; see Supplementary
information Table S1). DNA sequences were obtained from samples collected from the wild
following ethical guidelines (Sikes, Gannon & the Animal Care and Use Committee of the
American Society of Mammalogists, 2011), or from museums, and from published GenBank
data (including AB175132: Ohdachi *et al.*, 2006; AJ535393 – AJ535457: Mascheretti *et al*.,
2003; GQ272492 – GQ272518: Vega *et al*., 2010a; GQ494305 – GQ494350: Vega *et al*.,
2010b; and JF510376 – JF510321: McDevitt *et al*., 2011). In addition, four *MT-CYB*
sequences of *S. volnuchini*, which was used as an outgroup (Fumagalli *et al*., 1999), were
incorporated into the analysis (including AJ535458: Mascheretti *et al*., 2003).

Genomic DNA from wild and museum samples was extracted using a commercial kit
(Qiagen). Partial (1110 bp) *MT-CYB* sequences were obtained by PCR using two primer
pairs that amplified approximately 700 bp of overlapping fragments, or using five primer pairs
(for museum samples with highly degraded DNA) that amplified approximately 250 bp of
overlapping fragments (Vega *et al*., 2010a). PCR amplification was performed in a 50 μl final
volume: 1X Buffer, 1 μM each primer, 1 μM dNTP’s, 3 mM MgCl₂ and 0.5 U Platinum Taq
Polymerase (Invitrogen), with cycling conditions: 94°C for 4 min, 40 cycles at 94°C for 30 s,
55°C for 30 s and 72°C for 45 s, and a final elongation step at 72°C for 7 min. Purification of
PCR products was done with a commercial kit (Qiagen) and sequenced (Macrogen and
Cornell University Core Laboratories Center).
Phylogenetic analysis

DNA sequences were edited by eye in BioEdit version 7.0.9.0 (Hall, 1999), contigs were made from forward and reverse sequences also in BioEdit, and sequences were aligned using ClustalX version 2.0 (Larkin et al., 2007). A haplotype data file was obtained using DnaSP version 5.10.1 (Librado & Rozas, 2009). Newly obtained haplotypes were deposited in GenBank (Accession Numbers: MN840358 - MN840484, Supplementary information Table S1).

The model of evolution that best fitted the molecular data (haplotypes) was searched using jModelTest version 2.1.10 (Darriba et al., 2012) using the Bayesian Information Criterion value. The substitution model supported was the General Time Reversible (GTR) with specified substitution types (A–C=0.4250, A–G=23.5124, A–T=1.6091, C–G=1.8671, C–T=17.2314, G–T=1.0000), proportion of invariable sites (0.6044), gamma shape parameter (0.2816) and nucleotide frequencies (A=0.2777, C=0.3076, G=0.1416, T=0.2731).

The phylogenetic relationships among MT-CYB haplotypes of S. minutus were inferred by Bayesian analysis and by generating a parsimony phylogenetic network. The Bayesian analysis was done using MrBayes version 3.2.7 (Ronquist et al., 2012) with two independent runs (10 million generations and 5 chains each), a sampling frequency every 1000 generations and temperature of 0.1 for the heated chain, and checking for convergence in Tracer version 1.7.1 (Rambaut et al., 2018). Trees were summarized after a burn-in value of 2500 to obtain the posterior probabilities of each phylogenetic branch. The main phylogenetic groups (phylogroups) were identified based on monophyly of the haplotypes, and were named based on the geographical origin of the samples. The phylogenetic network was done using PopART version 1.7 (http://popart.otago.ac.nz) implementing a median-joining algorithm.

Sequence polymorphism indices and diversity values, including the number of haplotypes (H), polymorphic (segregating) sites (S) and parsimony informative sites (P), haplotype diversity (Hd), nucleotide diversity (π), and average number of nucleotide differences (k), were estimated using DnaSP. This was done for the whole data set (ingroup),
for the main phylogroups, and also for other relevant geographic groups, including island populations and continental samples.

Population genetic structure

Pairwise genetic differentiation values ($F_{ST}$) between all pairs of phylogroups and other relevant geographic groups, and an Analysis of Molecular Variance (AMOVA) were calculated using Arlequin version 3.11 (Excoffier et al., 2005). Ten thousand nonparametric permutations were performed to generate a random distribution to test the significance of the pairwise $F_{ST}$ values and covariance components of the AMOVA, and $\alpha = 0.05$ was set as the threshold for statistical significance.

With samples assigned to phylogroups and with the samples' geographical coordinates, the geographic midpoints (i.e. mid-geographic location between two or more coordinates) of the phylogroups were calculated using the Geographic Midpoint Calculator (available at http://www.geomidpoint.com/). The geographic midpoints were used to obtain the pairwise geographic distances among phylogroups with the Geographic Distance Matrix Calculator version 1.2.3 (by PJ Ersts, available at http://biodiversityinformatics.amnh.org/open_source/gdmg). A Mantel test was used to evaluate the relationship between matrices of pairwise geographic distances and genetic differentiation values (Slatkin’s linearised pairwise $F_{ST}$ as $D = F_{ST}/(1 - F_{ST})$; Slatkin, 1995). Despite criticisms, the Mantel test is still a widely used and can be a powerful statistical approach to analyse sequence data to test evolutionary hypotheses (Diniz-Filho et al., 2013). Due to the very low (or absence of) genetic variation in the Orkney islands, DNA sequences originating from there were pooled to avoid issues with pairwise $F_{ST}$ calculations.

Geographic barriers were computed using Barrier version 2.2 (Manni et al., 2004). This approach implements Monmonier’s maximum difference algorithm to find edges (boundaries) on a Voronoi tessellation associated with the highest rate of change in genetic distances among samples interconnected by a geometric network (i.e. Delaunay triangulation) (Manni et al., 2004). A barrier highlights the geographic areas where a genetic
discontinuity is found, and where samples on each side of the barrier are genetically more
similar than samples taken on different sides of the boundary. Pairwise genetic distances
were estimated using continental samples only, limiting the data set in the geometric network
calculation to one sample per locality, and computing a maximum of 10 barriers.

Historical demography

A strict molecular clock was compared to the uncorrelated lognormal relaxed molecular clock
(Drummond et al., 2006). Coalescent constant population size and Bayesian skyline
demographic models (Drummond et al., 2005) were compared to identify the best-fitting
pattern of changes in the pygmy shrew population. For model selection, path sampling and
stepping-stone sampling (Baele et al., 2013), based on four independent MCMC chains
(1000 steps of 100,000 generations each, following a 10 million generations burn-in period),
were used for calculating the log Marginal Likelihoods Estimates (MLEs) for each model.
MLEs were used to calculate Bayes Factors (BFs) for each comparison between tested
models to determine the best-fitting one (Kass & Raftery, 1995). The best-fitting models were
then used to estimate the Time of divergence from the Most Recent Common Ancestor
(TMRCA) and Bayesian Skyline Plots (BSP) (see below). The 95% Highest Posterior
Density (HPD) was included in the TMRCA and BSP estimations.

TMRCAs for the ingroup (all S. minutus samples) and the phylogroups were
estimated using BEAST version 2.5.2 (Bouckaert et al., 2014). The following prior
assumptions were: random starting tree, monophyletic groups (for the ingroup and the Irish
phylogroups) (Drummond et al., 2006) to calculate the evolutionary rate, and the GTR
substitution model with four categories, gamma = 0.9680 and proportion of invariable sites =
0.4680 (from jModelTest using the full data set). The oldest record of S. minutus has been
found in Podlesice and Mala Cave, Poland dated between 5.3 and 3.6 MYA (Early Pliocene;
Mammal Neogene 14) (Rzebik-Kowalska, 1998). Using this fossil information, a calibration
point for the ingroup was set at 4.45 MYA (SD = 0.5 MY; 5.27 – 3.63 MYA) with a normal
prior distribution. Due to the absence of dated fossils of pygmy shrews that can be assigned
specifically to the main phylogroups, a second calibration was set for the node age of the Irish lineage at 0.006 MYA (SD = 0.0005 MYA; 0.00682 – 0.00518 MYA). This secondary calibration point, derived from a previous analysis and applied to our data set, was based on the inferred colonisation time of Ireland by *S. minutus* in the Neolithic using multiple genetic markers and fossil data (McDevitt *et al.*, 2009, 2011). The trace files were analysed in Tracer, the tree information from the four runs were combined and resampled at a lower frequency (for a total of 10,000 trees) using LogCombiner, and the information was summarized using TreeAnnotator selecting Maximum clade credibility tree and median heights. The phylogenetic tree showing the TMRCAs was created using FigTree version 1.4.4 ([http://tree.bio.ed.ac.uk/software/figtree/](http://tree.bio.ed.ac.uk/software/figtree/)) with median and 95% HPD values based on those 10,000 trees.

Genetic evidence of population expansion for the phylogroups, island populations and continental samples was investigated using the R² test of neutrality (Ramos-Onsins & Rozas, 2002), based on the difference of the number of singleton mutations and the average number of nucleotide differences, and Fu's Fs (Fu, 1997), a statistic based on the infinite-site model without recombination that shows large negative Fs values when there has been a demographic population expansion. Both population expansion tests were carried out in DnaSP using coalescent simulations for testing significance (10,000 replicates).

Mismatch distributions (i.e. the distribution of the number of differences between pairs of haplotypes) were estimated for the phylogroups (and where N ≥ 10) to compare the demography of the populations with the expectations of a sudden population expansion model (Rogers & Harpending, 1992). For the phylogroups and continental samples that showed a unimodal mismatch distribution and significant population expansion, the time since the population expansion (*t*) was calculated as *t* = *r*/2*u*, where *r* (tau) is the mode for the unimodal mismatch distribution, and *u* is the cumulative (across the sequence) probability of substitution (Schenekar & Weiss, 2011). The calculations were done using the MS Excel Mismatch Calculator (Schenekar & Weiss, 2011) with sequence length = 1110 bp, generation time = 1 year (Hutterer *et al.*, 2016), substitutions per site per million years.
RESULTS

Phylogenetic analysis

For the complete *S. minutus* data set (*N* = 671) (Fig. 1B), there were 424 haplotypes with 390 polymorphic sites of which 277 were parsimony informative (Table 1). We report 160 newly sequenced specimens of *S. minutus* from the Iberian (4) and Balkan (19) peninsulas and from Central and Northern Europe (137) from which 127 were new haplotypes. Also, there were three new sequences and haplotypes of *S. volnuchini*, from which two were from Turkey and one from the Crimean Peninsula.

The Bayesian phylogenetic analysis showed *S. minutus* as a monophyletic group and revealed six distinct lineages corresponding to their geographical origin (i.e. phylogroups) supported by high posterior probabilities (Fig. 2A). Samples from the Mediterranean peninsulas clustered in three distinct phylogroups, namely the Iberian, Italian and Balkan phylogroups. The Iberian group was represented with few DNA sequences (*N* = 6). It was geographically restricted to the Iberian Peninsula and included samples from Rascafría, Central Spain (Sierra de Guadarrama) and Picos de Europa, Northern Spain. The Italian phylogroup (*N* = 26) was mostly restricted to the north-central regions of the Italian peninsula; it included samples from the Apennines and the Alps in Italy, but also from Switzerland, Slovenia, Southern and Eastern France near the border with Italy, Czech Republic and Germany. The Balkan phylogroup (*N* = 22) included samples mostly from the Balkan Peninsula and a few from further north in Central Europe. This phylogroup showed a
weak north/south subdivision, with one clade containing samples from Switzerland, Austria, Slovakia, Czech Republic, Hungary and Montenegro, another clade containing samples from Serbia, Bosnia and Herzegovina and North Macedonia, plus other ungrouped basal samples from Montenegro, North Macedonia, Serbia and Turkey (East Thrace, Southeast Europe).

There was also a well-supported and geographically widespread Western phylogroup \((N = 283)\), which included samples from northern Spain (Cantabrian Mountain Range), Southern France and Andorra (i.e. the Pyrenees), western and central France (including Belle-Île), Ireland, the Orkney Islands, and western mainland Britain and offshore islands on the western coast of mainland Britain. Samples from Ireland formed an internal monophyletic lineage (i.e. the Irish phylogroup, \(N = 94\)) within the Western phylogroup. Notably, two samples from Navarra in northern Spain (ESNa0861 and ESNa1131; Accession Number JF510331) shared haplotypes with samples from Ireland (Hap_64). A monophyletic South Italian phylogroup \((N = 4)\) was most closely related to the Western phylogroup than to the Italian phylogroup, and was geographically restricted to La Sila, Calabria in Southern Italy.

Samples from northern and central Europe and Siberia, namely the Northern phylogroup \((N = 330)\), formed the most geographically widespread lineage and included samples ranging from Central France and Britain (excluding those within the Western phylogroup), across Central and Northern Europe to Lake Baikal in Siberia, but did not include samples from Southern Europe. Samples from mainland Britain belonging to the Northern phylogroup did not form an internal monophyletic cluster.

The phylogenetic network had a complex structure (Fig. 2B), but the haplotypes clustered into the same phylogroups detected with Bayesian phylogenetics and were distantly related from each other (> 10 mutational steps). The Western phylogroup had a star-like pattern and showed three most internal haplotypes; notably, Hap_61 was found in the Pyrenees with other Western haplotypes directly connected to it, Hap_94 was found on islands of the western coast of Scotland (Arran and Mull) with other Scottish and continental Western haplotypes directly connected to it, and Hap_64 included samples from Northern Spain and Ireland with other Irish haplotypes connected to it. The Northern phylogroup
showed a star-like pattern with many reticulations and three most internal haplotypes separated from each other by few mutational steps. There was a weak geographical subdivision within the Northern phylogroup, where samples from Siberia, Eastern and Northern Europe were derived or most closely connected to samples from Central Ukraine (Hap_287), samples from Central Europe were derived or most closely connected to samples from The Netherlands (Hap_274), and all samples from Britain were derived or most closely connected to other samples from The Netherlands than to the other central haplotypes (Hap_90); however, the highly reticulated pattern of the inner haplotypes of the Northern phylogroup indicated that this geographical subdivision was weak.

Sequence polymorphism indices and diversity values for the phylogroups and other geographic groups are shown in Table 1. For the phylogroups, the haplotype diversity values were high (>90%), and the nucleotide diversity values were either half or almost half as much as the ingroup. Notably, the Northern phylogroup had the highest haplotype diversity values, followed by the Balkan phylogroup; however, the Balkan phylogroup had the highest nucleotide diversity values. The Irish phylogroup, which clustered within the Western phylogroup, showed slightly lower haplotype diversity than any other phylogroups.

The continental groups (Northern continental and Western continental) showed equivalent DNA polymorphism values as the main phylogroups, but the island groups showed different levels of DNA polymorphism (Table 1). There was low DNA polymorphism in islands of the Orkney Archipelago, with only 11 haplotypes in all Orkney Islands combined (N = 119), but all haplotypes were unique to these islands. There were eight haplotypes in Orkney Mainland (N = 44), from which seven were unique to this island (the largest island of the archipelago), there were two unique haplotypes in Orkney South Ronaldsay (N = 40), and there was only one haplotype in Orkney Westray (N = 33) also present in Orkney Hoy (N = 2) and Orkney Mainland. There were five haplotypes in Belle-Île (N = 5), and only one was present in the continent also belonging to the Western phylogroup. The British group (N = 91) showed high haplotype diversity but moderate nucleotide diversity values and had 80 haplotypes from which 77 were unique haplotypes not found elsewhere.
Population genetic structure

The highest pairwise differentiation values were found between some southern phylogroups and island groups, while the lowest values were between phylogroups and islands groups that clustered within them (Supplementary information Table S2). There was higher percentage of variation among (73.5 %) than within (26.5 %) groups, and there was a significant population differentiation ($F_{ST} = 0.7349, P < 0.0001$). The Mantel test showed a nonsignificant relationship between pairwise geographic and genetic distances based on Slatkin’s linearised $F_{ST}$ ($R^2 = 0.0095, P = 0.2935$) (Supplementary information Fig. S1).

The barriers identified using the computational geometry approach reflected the genetic differentiation between S. minutus and S. volnuchini, and among the phylogroups within S. minutus (Fig. 1C). The first barrier separated S. minutus from S. volnichini. The nine following barriers coincided with the location of mountain ranges, including a barrier located in the north of the Balkan Peninsula, in the Alps and in the Pyrenees, which reflected the genetic subdivisions and lineages in S. minutus.

Historical demography

Comparison of BFs for each model indicated the Bayesian skyline demographic model as the best-fitting one (BF = 391), and the strict molecular clock was better than the uncorrelated lognormal relaxed molecular clock (BF = 23). The MLEs for the constant population size and Bayesian skyline demographic models using the strict molecular clock were -10960 and -10569, while using the uncorrelated lognormal relaxed molecular clock were -10907 and -10592, respectively. Therefore, the strict clock and Bayesian skyline demographic model were selected as the best-fitting according to BFs. The effective sample size (ESS) for all values was higher than 200.

All branches of the Bayesian genealogy (Fig. 3, Table 2) were well-supported (posterior probabilities PP ≥ 0.97), except for the clade containing all phylogroups excluding Iberian (PP = 0.05). Molecular dating analysis revealed that the ingroup and outgroup
separated approximately 83.4 KYA, with lower and upper 95% HPD limits between 59.7 and 110.2 KYA (Fig. 3, Table 2). The diversification of *S. minutus* occurred approximately 31.8 KYA (95% HPD: 21.8 – 40.5 KYA) with the formation of the Iberian phylogroup, followed by the formation of the Balkan group 29.6 KYA (95% HPD: 21.8 – 40.5 KYA), while in Western, Central and Northern Europe, *S. minutus* continued its diversification with the Northern phylogroup forming 24.1 KYA (95% HPD: 16.4 – 33.1 KYA), the Italian phylogroup forming 15.3 KYA (95% HPD: 10.7 – 21.5 KYA), the South Italian phylogroup forming 12.8 KYA (95% HPD: 8.5 – 17.8 KYA), and the Western phylogroup forming 9.3 KYA (95% HPD: 6.7 – 12.6 KYA) (Fig. 3, Table 2). The TMRCA of the Balkan phylogroup was the earliest, dated back to 15.5 KYA (95% HPD: 9.7-22.7 KYA), followed by the Northern phylogroup, dated back to 11.8 KYA (95% HPD = 7.7-16.8 KYA), while the rest of the main phylogroups had TMRCAs dated approximately to about 6 and 9 KYA (Fig. 3, Table 2). Within the Western phylogroup, the TMRCA for the Irish clade dated back to 5.9 KYA (95% HPD: 4.9 – 6.9 KYA).

The population expansion tests (*R*<sub>2</sub> and Fu's *Fs*) showed significant departures from neutrality for the ingroup and several other phylogroups, except for the Balkan, Iberian and South Italian (Table 2). The population expansions were not an effect of the island samples belonging to these phylogroups, and continental samples analysed separately also demonstrated a similar pattern (Table 2). For the island groups, only the Irish and British groups showed signatures of recent population expansions (Table 2).

The mismatch distributions varied significantly among the phylogroups (Fig. 4A; Supplementary information Fig. S2). The ingroup showed a bimodal mismatch distribution, which reflected the pairwise comparisons within and among phylogroups in *S. minutus*. The Northern (and Northern continental), Italian, Western (and Western continental) and Irish phylogroups all had distinctly unimodal distributions with an almost perfect fit between observed and expected pairwise differences of a sudden population expansion model. All population expansions for the phylogroups were dated to the Holocene; the Italian and Northern phylogroups had the oldest times of expansion (>8.0 KYA), while the Irish showed a relatively recent population expansion dated to 1.6 KYA.
The BSP obtained for three phylogroups (Northern, Western and Irish) suggested that demographic expansions of these populations started approximately 5.0 KYA (Fig. 4B). BSP calculation for the Italian phylogroup indicated an even earlier demographic expansion (approximately 5.5 KYA) (Fig. 4B).

DISCUSSION

Quaternary refugia represent the geographical regions that species inhabit during periods of glacial or interglacial cycles when there is the maximum contraction in geographical range (Stewart et al., 2009). There is support for both southern (Taberlet et al., 1998; Hewitt, 2000) and northern glacial European refugia (Bilton et al., 1998; Stewart & Lister, 2001; Kotlík et al., 2006; Provan & Bennett 2008; Flojgaard et al., 2009; Vega et al., 2010a, b). Rather than polarising the biogeographic patterns into southern and northern refugia (Tzedakis et al., 2013), the paradigms of postglacial colonisation in Europe (Hewitt, 2000) can be improved with the acceptance of southern hotspots of diversification without northward colonisation (Bilton et al., 1998) and the concept of refugia-within-refugia (Gómez & Lunt, 2007), as well as with the findings of northern glacial refugia (Stewart & Lister 2001; Provan & Bennett, 2008; Stewart et al., 2009), to reflect the evolutionary processes across varied topographical areas that have shaped genetic diversity. The statistical phylogeographic results obtained here contribute to the understanding of the phylogeographic patterns and processes during the Quaternary glaciations that shaped the European biota, and provide further evidence to the emerging pattern of complex biogeographical histories in Europe (Pedreschi et al., 2019).

Sorex minutus phylogeography

The significant genetic structure among phylogroups defined in this study illustrate the complex history of European colonisation, isolation and diversification of S. minutus during the Pleistocene and Holocene, and is not a simple case of isolation by distance and colonisation of Northern and Central Europe from expanding populations from the south.
While the southern phylogroups, including the Iberian, Balkan, Italian and South Italian, were mostly restricted to the Southern European peninsulas (consistent with the traditional southern glacial refugia), the Northern and Western phylogroups were widespread geographically and were found north of the Mediterranean peninsulas, consistent with previous studies with fewer samples (Bilton et al., 1998; Mascheretti et al., 2003; Vega et al. 2010a, b) and with different molecular markers (McDevitt et al., 2010).

The hypothesis of northern refugia is further supported by palaeontological and palynological evidence for other temperate and boreal species (Willis et al., 2000; Willis & van Andel, 2004; Magri et al., 2006; Sommer & Nadachowski, 2006), as well as many phylogeographic studies in small mammals, including the field vole M. agrestis (Jaarola & Searle, 2002; Herman et al., 2019), bank vole M. glareolus (Deffontaine et al., 2005; Kotlík et al., 2006; Wójcik et al., 2010), root vole M. oeconomus (Brunhoff et al., 2003), common vole M. arvalis (Heckel et al., 2005; Stojak et al., 2016), common shrew S. araneus (Bilton et al., 1998; Yannic et al., 2008) and weasels Mustela nivalis (McDevitt et al., 2012). For several small mammals, including S. minutus, suitable climatic conditions at the LGM could have been widespread across Central and Eastern Europe (Fløjgaard et al., 2009; Vega et al., 2010b; McDevitt et al. 2012; Stojak et al., 2019).

Until recently, it was unclear which species of Sorex inhabit Crimea. According to Zagorodniuk (1996) it could be S. (minutus) dahli [mentioned in Hutterer (2005) as a synonym of Sorex volnuchini (dahli)], and Zaitsev et al. (2014) and Hutterer et al. (2016) indicated the presence of S. minutus in mainland Ukraine and in Crimea. Hutterer (2005) mentioned that S. volnuchini might be present in Crimea, but in Hutterer et al. (2016) S. volnuchini is only found in southern Russia and Caucasus States, Turkey and northern Iran. Our research demonstrated that S. volnuchini may be present in the southern region of Crimea (based on one MT-CYB sequence), while S. minutus is present in mainland Ukraine and in the northern region of Crimea, but further sampling in this region is needed.

The finding of two phylogroups in the Iberian peninsula (i.e. Iberian and Western phylogroups) and two in the Italian peninsula (i.e. Italian and South Italian phylogroups),
support the hypothesis of microevolutionary processes shaping the genetic diversity and
structure within the Mediterranean peninsulas. In the Iberian Peninsula, the topography of
the region with east-west mountain ranges and other high ground (over 1000 m a.s.l.), large
rivers (which could act as barriers to dispersal), and the distinct seasonal precipitation and
vegetation types (O’Regan, 2008), must have played an important role in the colonisation of
the region and the genetic differentiation of populations. McDevitt et al. (2010) proposed that
the Western phylogroup could have originated in the Dordogne region based on a limited
number of samples from France, but the presence of this phylogroup in northern Iberia and
the central position of Hap_61 (Pyrenees) could mean that an Iberian origin is possible
instead. A similar process could explain the presence of the two phylogroups in the Italian
peninsula (i.e. Italian and South Italian). The genetic differentiation of the South Italian
phylogroup, further supported by morphological data (Vega et al., 2010a, 2016), could be
due to the unique geography of Southern Italy consisting of mountain massifs of Pollino, La
Sila and Aspromonte separated by lowland areas, which from the Pliocene to the end of the
Middle Pleistocene, at times of high sea level, were islands in a chain (Malatesta, 1985;
Caloi et al., 1989; Bonardi et al., 2001; Bonfiglio et al., 2002). The patterns of differentiation
within refugial areas were concordant with the ‘refugia-within-refugia’ concept widely
recognized for the Iberian Peninsula (Gómez & Lunt, 2007; Abellán & Svenning, 2019) and
similar to microrefugia in the Balkans (Kryštufek et al., 2007). For the Italian peninsula, a
comparable ‘refugia-within-refugia’ pattern was found in several species (Amori et al., 2008;
Canestrelli et al., 2008; Castiglia et al., 2008; Vega et al., 2010a, 2016; Senczuk et al., 2017;
Bisconti et al., 2018).

The genetic similarity between the Western and South Italian phylogroups indicates a
common history and it can be hypothesised that their common ancestor was more
widespread throughout the Italian peninsula, probably displaced later by the Italian lineage in
the Apennines and Western Alps. A similar scenario has been proposed for the water shrew
Neomys fodiens (Castiglia et al., 2007), Alpine salamander Salamandra salamandra
(Steinfartz et al., 2000), black pine Pinus nigra (Afzal-Rafii & Dodd, 2007) and green lizard
Lacerta bilineata bilineata (Böhme et al., 2007), which showed closely related South Italian and Western phylogroups most closely related to each other than to a North-Central Italian lineage.

The phylogeographic patterns found here were further supported by the determination of barriers that coincided with mountain ranges located on the north of the Iberian, Italian and Balkan peninsulas. Contact zones among phylogroups (i.e. localities where at least two MT-CYB phylogroups were present) were detected at the northern extremes of the southern peninsulas. During the LGM, glaciers covered most of the Alpine (Buoncristiani & Campy, 2004) and Pyrenean mountain ranges (Calvet, 2004), while glaciers in the Carpathians (Reuther et al., 2007) and in the Balkan Peninsula (Hughes et al., 2006) were found > 1,000 m a.s.l. When climate ameliorated and suitable habitat became available, pygmy shrew populations belonging to different phylogroups on different sides of the mountain ranges could have expanded and colonised previously glaciated areas thus forming the observed contact zones. Moreover, the widespread distribution and absence of phylogeographic structure of the Northern phylogroup could be explained by the apparent absence of major geographical barriers across Central and Northern Europe, and recolonization from northern refugia. Similarly, pygmy shrews belonging to the Western and Northern phylogroups could have quickly colonised mainland Britain across a land connection to continental Europe (i.e. Doggerland; Gaffney et al., 2007), resulting in the genetic similarities observed between the British Isles and continental Europe.

Sorex minutus demography

The oldest fossils assigned to S. minutus were found in Podlesice and Mała Cave, Poland dated to the Early Pliocene between 4 and 5.3 MYA (Rzebik-Kowalska, 1998). An early widespread colonisation of Europe by S. minutus might have been possible because it was probably one of the first species of the genus Sorex in the continent (Rzebik-Kowalska, 1998, 2008). The Bayesian analysis revealed, however, more recent diversification events, with TMRCAs for the ingroup and the phylogroups in continental Europe between the Upper
Pleistocene and Lower Holocene. This is consistent with recent studies on field vole 
(Microtus agrestis) phylogeography in Europe (Herman & Searle, 2011; Herman et al., 2014) 
which demonstrated the importance of the Younger Dryas (11.7-12.9 KYA) glacial re-
advance for the diversification within this species. Similar colonisation scenarios and 
divergence before the LGM from Eastern to Western Europe have been proposed for other 
species, including the common vole Microtus arvalis (Heckel et al., 2005; Stojak et al., 
2016), the bank vole Clethrionomys glareolus (Deffontaine et al., 2005; Kotlík et al., 2006; 
Wójcik et al., 2010), and the root vole M. oeconomus (Brunhoff et al., 2003).

The population expansion signatures for the Northern and Western phylogroups, 
star-like patterns in phylogenetic networks and population expansion times support recent 
and quick colonisation events of central and northern Europe, and appear to reflect 
responses to postglacial climate warming. The Western lineage was restricted to Central, 
Western and South-Eastern France and North-Western Spain in continental Europe, but it 
was the only lineage found in Ireland and several islands off the west and north coasts of 
Britain. The region of the Dordogne in South-Western France was situated outside the LGM 
permafrost area and has temperate mammal fossil records dated to the end of the LGM; 
therefore, it has been suggested as another likely northern refugium (Sommer & 
Nadachowski, 2006; McDevitt et al., 2010) where the Western lineage could have persisted 
and recolonised Western and Central France after the LGM. But as stated above, an Iberian 
origin for this phylogroup is also possible. However, SDM studies showed that suitable 
climatic conditions during the LGM for S. minutus and other temperate small mammal 
species could have been more continuous and present further north (Fløjgaard et al., 2009; 
Vega et al., 2010b), which could explain its widespread distribution in Western Europe and 
its presence in Britain. According to BSP results, it is plausible that Northern and Western 
phylogroups spread across Europe after the Younger Dryas. The British (island) group, 
belonging to the Northern phylogroup, showed a significant signature of population 
expansion. This expansion could have selectively displaced pygmy shrew populations of the
Western lineage, which still remain in uplands and islands in the periphery to the north, west and south of Britain forming a ‘Celtic fringe’ (Searle et al., 2009).

The widespread Italian lineage may be presumed to derive from a glacial refugium located somewhere within the vicinity of the Apennine mountain chain. A significant population expansion signature demonstrates that the Italian phylogroup went through a recent expansion phase, calculated in BSP for about 5.5 KYA. Contrastingly, the lack of a population expansion signature, the high nucleotide and haplotype diversities, and the highly divergent sequences showing a weak north/south subdivision of the Balkan phylogroup warrants further attention. The Balkans is a European hotspot for biodiversity given its environmental stability, topographic and climatic diversity and occasional connectedness with Asia Minor (Kryštufek & Reed, 2004; Kryštufek et al., 2007, 2009; Bužan et al., 2010), and it could be expected that some of these factors shaped the genetic diversity of the Balkan lineage there. Similarly, the lack of significant population expansion values for the Iberian lineage may relate to historical stable population sizes; however, the sample size was low and this result should be taken with caution.

Further considerations and implications

The comparison of the results obtained here with those elsewhere shows an emerging pattern of glacial refugia in Mediterranean peninsulas and further north in Central Europe for several species.

Although S. minutus is considered as a least concern species by the IUCN (Hutterer et al., 2016), the distinct phylogroups deserve more attention than this implies. Genetic diversity is considered an important aspect of global biodiversity (McNeely et al., 1990), and local and/or country-based conservation efforts are highly valued (for example, in Britain and Ireland the pygmy shrew is protected by law). The refugial areas in Southern Europe are often found in mountain ranges at the low-latitude margins of the present-day distribution ranges of species and are most likely to contain rear-edge populations where selection for local adaptations could have resulted in the evolution of distinct ecotypes (Cook, 1961;
Hampe & Petit, 2005). Rear-edge populations, including the southern lineages of *S. minutus*, deserve further investigation and should be regarded for conservation because they are important to determine the responses of species to modern climate change (Petit *et al.*, 2003; Hampe & Petit, 2005).

In conclusion, the Eurasian pygmy shrew *Sorex minutus* is a good model for understanding biological diversity, colonisation patterns and the effects of past climate change on biological diversity. There is a mosaic of genetic lineages across continental Europe, characterised by different demographic histories and natural colonisation patterns, while island populations are characterised by recent natural and human-mediated colonisations. This study has notably expanded previous findings on *S. minutus*, with a more precise statistical phylogeographic analysis of the genetic variability and structure, colonisation routes, geographical barriers and historical demography across Europe.

Specifically, we provided new data from the Iberian and Balkan peninsulas, and from Central and Eastern Europe (Poland, Ukraine and Russia), important for understanding postglacial events. *Sorex minutus* is not an easy species to obtain in large numbers, and the sampling described here represents a very substantial effort. However, it is a species that is unusually widespread and genetically subdivided and therefore can inform better than almost any other about the relative importance of southern and northern glacial refugia.

**ACKNOWLEDGEMENTS**

Specimens and species records of *Sorex minutus* were made available by several museums and we acknowledge the help of the curators from the following institutions: Département d’écologie et évolution (Université de Lausanne, Switzerland), Univerza na Primorskem and Research Centre of Koper (Slovenia), Natuurhistorisch Museum (Rotterdam, Netherlands), Dipartimento di Ecologia (Università della Calabria, Italy), Museo di Anatomia Comparata and Museo di Zoologia “La Sapienza” (Università di Roma, Italy), Natuurmuseum Brabant (Tilburg, Netherlands) and Mammal Research Institute PAS (Bialowieża, Poland). We are very grateful for the tissue samples provided by Glenn Yannic, Jacques Hausser, Jan Zima,
Fríða Jóhannesdóttir, Holger Bruns, Peter Borkenhagen, Natália Martinková, Elena Gladilina, Pavel Goldin and Petr Kotlík. We thank Ricardo Castiglia and two anonymous reviewers for their helpful comments.

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. 


Bonardi G, Cavazza W, Perrone V, Rossi S. 2001. Calabria-Peloritani terrane and
Northern Ionian Sea. In: Vai GB, Martini IP, eds. Anatomy of an orogen: The


Drummond AJ. 2014. BEAST 2: a software platform for Bayesian evolutionary
analysis. PLOS Computational Biology 10: e1003537.

phylogeography of the root vole (Microtus oeconomus): implications for late

Buoncristiani J-F. Campy M. 2004. Palaeogeography of the last two glacial episodes in
Extent and chronology Part I: Europe. London: Developments in Quaternary Science
2: 101-110.

Bužan EV, Förster DW, Searle JB, Kryštufek B. 2010. A new cytochrome b phylogroup of
the common vole (Microtus arvalis) endemic to the Balkans and its implications for
the evolutionary history of the species. Biological Journal of the Linnean Society 100:
788-796.

il Quaternario. Atti Accademia Peloritana dei Pericolanti Classe I, Scienze
Matematiche Fisiche e Naturali 67: 261-278.

Quaternary glaciations – Extent and chronology Part I: Europe. London:
Developments in Quaternary Science 2: 119-128.

of the Apennine endemic stream frog, Rana italic; insights on the Pleistocene


implications for colonization history. Biological Journal of the Linnean Society 97: 918-927.


O'Regan HJ. 2008. The Iberian Peninsula – corridor or cul-de-sac? Mammalian faunal change and possible routes of dispersal in the last 2 million years. Quaternary Science Reviews 27: 2136-2144.


Stojak J, McDevitt AD, Herman JS, Searle JB, Wójcik JM. 2015. Post-glacial colonization of eastern Europe from the Carpathian refugium: evidence from mitochondrial DNA of


SUPPORTING INFORMATION

Table S1. *Sorex minutus* dataset and sample information

Table S2. Pairwise geographic distances (in Km, below diagonal) and genetic differentiation (Slatkin's $F_{ST}$, above diagonal) among *Sorex minutus* phylogroups and other geographic groups

Figure S1. Correlogram of pairwise geographic and genetic distances among *Sorex minutus* phylogroups and other geographic groups.

Figure S2. Mismatch distributions of *Sorex minutus* phylogroups and other geographic groups.
Table 1. DNA sequence polymorphism in *Sorex minutus* phylogroups and other geographic groups

<table>
<thead>
<tr>
<th>Phylogroups</th>
<th>N</th>
<th>S</th>
<th>P</th>
<th>H</th>
<th>Hd</th>
<th>Hd (SD)</th>
<th>π</th>
<th>π (SD)</th>
<th>k</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ingroup</td>
<td>671</td>
<td>390</td>
<td>277</td>
<td>424</td>
<td>0.9899</td>
<td>0.0015</td>
<td>0.0143</td>
<td>0.0000</td>
<td>15.8670</td>
</tr>
<tr>
<td>Italian</td>
<td>26</td>
<td>51</td>
<td>19</td>
<td>18</td>
<td>0.9600</td>
<td>0.0230</td>
<td>0.0061</td>
<td>0.0004</td>
<td>6.7720</td>
</tr>
<tr>
<td>South Italian</td>
<td>4</td>
<td>16</td>
<td>0</td>
<td>4</td>
<td>1.0000</td>
<td>0.1770</td>
<td>0.0072</td>
<td>0.0020</td>
<td>8.0000</td>
</tr>
<tr>
<td>Balkan</td>
<td>22</td>
<td>55</td>
<td>28</td>
<td>17</td>
<td>0.9610</td>
<td>0.0290</td>
<td>0.0097</td>
<td>0.0009</td>
<td>10.7970</td>
</tr>
<tr>
<td>Iberian</td>
<td>6</td>
<td>15</td>
<td>6</td>
<td>5</td>
<td>0.9330</td>
<td>0.1220</td>
<td>0.0058</td>
<td>0.0013</td>
<td>6.4000</td>
</tr>
<tr>
<td>Western</td>
<td>283</td>
<td>147</td>
<td>83</td>
<td>102</td>
<td>0.9458</td>
<td>0.0067</td>
<td>0.0049</td>
<td>0.0002</td>
<td>5.4400</td>
</tr>
<tr>
<td>Irish</td>
<td>94</td>
<td>53</td>
<td>21</td>
<td>42</td>
<td>0.8920</td>
<td>0.0270</td>
<td>0.0020</td>
<td>0.0002</td>
<td>2.2180</td>
</tr>
<tr>
<td>Northern</td>
<td>330</td>
<td>311</td>
<td>197</td>
<td>278</td>
<td>0.9984</td>
<td>0.0005</td>
<td>0.0065</td>
<td>0.0002</td>
<td>7.1840</td>
</tr>
<tr>
<td>Continental groups</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Western (Continental)</td>
<td>15</td>
<td>28</td>
<td>11</td>
<td>13</td>
<td>0.9810</td>
<td>0.0310</td>
<td>0.0050</td>
<td>0.0006</td>
<td>5.5430</td>
</tr>
<tr>
<td>Northern (Continental)</td>
<td>226</td>
<td>241</td>
<td>142</td>
<td>188</td>
<td>0.9978</td>
<td>0.0007</td>
<td>0.0062</td>
<td>0.0002</td>
<td>6.9300</td>
</tr>
<tr>
<td>Other island groups</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Orkney Islands (All)</td>
<td>119</td>
<td>17</td>
<td>13</td>
<td>11</td>
<td>0.7720</td>
<td>0.0210</td>
<td>0.0027</td>
<td>0.0001</td>
<td>3.0140</td>
</tr>
<tr>
<td>Orkney Mainland</td>
<td>44</td>
<td>9</td>
<td>7</td>
<td>8</td>
<td>0.7550</td>
<td>0.0550</td>
<td>0.0013</td>
<td>0.0002</td>
<td>1.4790</td>
</tr>
<tr>
<td>Orkney South Ronaldsay</td>
<td>40</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>0.1420</td>
<td>0.0710</td>
<td>0.0001</td>
<td>0.0001</td>
<td>0.1420</td>
</tr>
<tr>
<td>Orkney Westray</td>
<td>33</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
</tr>
<tr>
<td>Orkney Hoy</td>
<td>2</td>
<td>2</td>
<td>0</td>
<td>2</td>
<td>1.0000</td>
<td>0.5000</td>
<td>0.0018</td>
<td>0.0009</td>
<td>2.0000</td>
</tr>
<tr>
<td>Belle Île</td>
<td>5</td>
<td>9</td>
<td>3</td>
<td>5</td>
<td>1.0000</td>
<td>0.1260</td>
<td>0.0038</td>
<td>0.0010</td>
<td>4.2000</td>
</tr>
<tr>
<td>British</td>
<td>91</td>
<td>146</td>
<td>61</td>
<td>80</td>
<td>0.9960</td>
<td>0.0030</td>
<td>0.0055</td>
<td>0.0003</td>
<td>6.1210</td>
</tr>
</tbody>
</table>

*N* = Sample size; *S* = Number of polymorphic (segregating) sites; *P* = Parsimony informative sites; *H* = Number of haplotypes; *Hd* = Haplotype diversity; *SD* = Standard Deviation; *π* = Nucleotide diversity; *k* = Average number of nucleotide differences.
Table 2. Population expansion tests for *Sorex minutus* phylogroups and other geographic groups

<table>
<thead>
<tr>
<th>Phylogroups</th>
<th>$R_2$</th>
<th>$P$-value</th>
<th>$Fs$</th>
<th>$P$-value</th>
<th>$\tau$ (in years)</th>
<th>$t$ (in KYA)</th>
<th>TMRCA (in KYA)</th>
<th>95% HPD (in KYA)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ingroup</td>
<td>0.0198</td>
<td><strong>0.0004</strong></td>
<td>-741.2620</td>
<td>***</td>
<td>7.8590</td>
<td>6425</td>
<td>31.8</td>
<td>22.0-43.1</td>
</tr>
<tr>
<td>Italian</td>
<td>0.0521</td>
<td><strong>0.0000</strong></td>
<td>-5.8766</td>
<td><strong>0.0152</strong></td>
<td>6.7720</td>
<td>5536</td>
<td>7.2</td>
<td>4.8-10.2</td>
</tr>
<tr>
<td>South Italian</td>
<td>0.1822</td>
<td>0.1658</td>
<td>0.0687</td>
<td>0.2975</td>
<td>5.6340</td>
<td>-</td>
<td>7.7</td>
<td>4.2-12.1</td>
</tr>
<tr>
<td>Balkan</td>
<td>0.0830</td>
<td>0.0542</td>
<td>-3.6701</td>
<td>0.0768</td>
<td>7.1500</td>
<td>-</td>
<td>15.5</td>
<td>9.7-22.7</td>
</tr>
<tr>
<td>Iberian</td>
<td>0.1462</td>
<td>0.0888</td>
<td>0.0731</td>
<td>0.4290</td>
<td>4.0100</td>
<td>-</td>
<td>6.2</td>
<td>3.9-10.0</td>
</tr>
<tr>
<td>Western</td>
<td>0.0175</td>
<td><strong>0.0004</strong></td>
<td>-114.6990</td>
<td>***</td>
<td>3.6660</td>
<td>2997</td>
<td>9.3</td>
<td>6.7-12.6</td>
</tr>
<tr>
<td>Irish</td>
<td>0.0187</td>
<td><strong>0.0000</strong></td>
<td>-52.5664</td>
<td>***</td>
<td>1.3040</td>
<td>1066</td>
<td>5.9</td>
<td>4.9-6.9</td>
</tr>
<tr>
<td>Northern</td>
<td>0.0105</td>
<td><strong>0.0000</strong></td>
<td>-663.4730</td>
<td>***</td>
<td>6.5390</td>
<td>5346</td>
<td>11.8</td>
<td>7.7-16.8</td>
</tr>
</tbody>
</table>

Continental groups

| Western (Continental)     | 0.0793| **0.0045**| -6.0342 | **0.0035** | 5.5430 | 4532 | - | - |
| Northern (Continental)    | 0.0128| **0.0000**| -386.4520 | *** | 5.8010 | 4742 | - | - |

Other island groups

| Orkney Islands (All)      | 0.0880| 0.5209 | 0.6044 | 0.6437 | 1.1740 | - | - | - |
| Orkney Mainland           | 0.0839| 0.2301 | -1.6879 | 0.1892 | 1.4790 | - | - | - |
| Orkney Hoy                | 0.5000| 1.0000 | NC | NC | 2.0000 | - | - | - |
| Orkney South Ronaldsay    | 0.0712| 0.1770 | -0.2182 | 0.4420 | 0.1420 | - | - | - |
| Orkney Westray            | NC | NC | NC | NC | NC | - | - | - |
| Belle Île                 | 0.1915| 0.2467 | -1.6330 | 0.0732 | 3.5500 | - | - | - |
| British                   | 0.0161| **0.0000**| -122.8550 | *** | 6.1210 | 5004 | - | - |

*R_2* = Ramos-Onsins and Rozas (2002) test of neutrality; $P$-value = *P*-values of expansion tests expected under neutrality (**$P < 0.001$**); $Fs$ = Demographic population expansion test (Fu 1997); $\tau$ = (2ut) The mode of a mismatch distribution; $t$ = Time of population expansion (for phylogroups with bi- or unimodal mismatch distributions); TMRCA = Time of divergence from the Most Recent Common Ancestor; 95% HPD = 95% Highest Posterior Density; KYA = Thousand Years Ago; NC = Not computable (not enough variation or samples)
FIGURE LEGENDS

Figure 1. A) Map of Eurasia showing the geographical distribution of the Eurasian pygmy shrew *Sorex minutus* (Hutterer *et al.*, 2016). B) Sample localities of *S. minutus* used for this study and divided into mitochondrially encoded cytochrome b (*MT-CYB*) phylogroups (symbols with a dot represent samples used for inferring geographic barriers). C) Geographic barriers (red lines) for *S. minutus*; the barriers (up to a maximum of 10) were inferred using Monmonier’s maximum difference algorithm which finds edges (boundaries) on the Voronoi tessellation (blue polygons) associated with the highest rate of change in genetic distances among a subset of continental samples (dots) interconnected with a Delaunay triangulation (green lines).

Figure 2. Phylogenetic reconstructions of the Eurasian pygmy shrew *Sorex minutus* using *MT-CYB* sequences. A) Bayesian phylogenetic tree (with posterior probabilities on branches) showing the phylogroups. B) Haplotype phylogenetic network with haplotypes represented as nodes and their evolutionary relationships represented by edges; relevant haplotypes named at the centre of star-like patterns.

Figure 3. Maximum Clade Credibility tree for 671 sequences of *Sorex minutus* from Europe and Siberia, annotated from 10,000 Bayesian genealogy sampling. Posterior probabilities of basal nodes indicate support for each lineage. Horizontal bars represent 95% HPD intervals for Time to Most Recent Common Ancestor (TMRCA) of each lineage (in thousand years ago, KYA).

Figure 4. Historical demography of the Eurasian pygmy shrew *Sorex minutus*. A) Mismatch distributions of groups with significant signatures of population expansion. B) Bayesian Skyline Plots (BSP) of phylogroups with significant signatures of population expansion. The
solid lines in BSP are median estimates and the shaded areas represent 95% Highest Probability Densities (confidence intervals).
FIGURES

Figure 1.
Figure 2.

A) 

Iberian
Balkan
Italian
South Italian
Irish
Western
Northern

B) 

Hap_61
Hap_94
Hap_64
Hap_287
Hap_274
Hap_90

Legend:
- Western
- Northern
- Iberian
- Balkan
- Italian
- South Italian
- Irish
- 10 samples
- 1 sample
Figure 3.
Figure 4.

A) Ingroup

B) Ingroup

Effective population size

Italian

Western

Irish

Northern
1 SUPPLEMENTARY FIGURES

2 Figure S1.

![Graph showing the relationship between pairwise Slatkin’s FST and geographic distance. The equation of the line is y = 0.0002x + 2.6928 and R² = 0.0095.]
Figure S2.
SUPPLEMENTARY TABLES

Table S1. Sorex minutus dataset and sample information. Contact corresponding author.
Table S2. Pairwise geographic distances (in Km, below diagonal) and genetic differentiation (Slatkin's $F_{ST}$, above diagonal) among *Sorex minutus* phylogroups and other geographic groups.

<table>
<thead>
<tr>
<th></th>
<th>Italian</th>
<th>South Italian</th>
<th>Balkan</th>
<th>Iberian</th>
<th>Belle Île</th>
<th>Britain</th>
<th>Northern (Continental)</th>
<th>Western (Continental)</th>
<th>Orkney Islands</th>
<th>Irish</th>
</tr>
</thead>
<tbody>
<tr>
<td>Italian</td>
<td>-</td>
<td>1.6558</td>
<td>3.0387</td>
<td>3.6919</td>
<td>1.5940</td>
<td>2.8852</td>
<td>2.3798</td>
<td>1.4534</td>
<td>2.8673</td>
<td>4.0090</td>
</tr>
<tr>
<td>South Italian</td>
<td>773.14</td>
<td>-</td>
<td>2.5113</td>
<td>3.3869</td>
<td>1.3569</td>
<td>2.8662</td>
<td>2.3204</td>
<td>1.1713</td>
<td>2.7079</td>
<td>4.6820</td>
</tr>
<tr>
<td>Balkan</td>
<td>547.27</td>
<td>628.96</td>
<td>-</td>
<td>1.9234</td>
<td>2.5617</td>
<td>3.1494</td>
<td>2.8093</td>
<td>2.7975</td>
<td>6.2456</td>
<td>7.0533</td>
</tr>
<tr>
<td>Iberian</td>
<td>1349.26</td>
<td>1768.56</td>
<td>1880.98</td>
<td>-</td>
<td>4.2650</td>
<td>3.1850</td>
<td>2.6191</td>
<td>3.8498</td>
<td>7.3804</td>
<td>10.8797</td>
</tr>
<tr>
<td>Belle Île</td>
<td>1162.82</td>
<td>1815.47</td>
<td>1701.44</td>
<td>640.58</td>
<td>-</td>
<td>2.7595</td>
<td>2.1790</td>
<td>0.3345</td>
<td>1.2148</td>
<td>2.4003</td>
</tr>
<tr>
<td>Britain</td>
<td>1347.34</td>
<td>2108.66</td>
<td>1795.12</td>
<td>1286.79</td>
<td>647.11</td>
<td>-</td>
<td>0.1449</td>
<td>2.6083</td>
<td>3.9225</td>
<td>4.7265</td>
</tr>
<tr>
<td>Northern (Continental)</td>
<td>1022.36</td>
<td>1488.84</td>
<td>863.37</td>
<td>2218.78</td>
<td>1788.28</td>
<td>1554.91</td>
<td>-</td>
<td>2.0767</td>
<td>2.7910</td>
<td>3.2035</td>
</tr>
<tr>
<td>Western (Continental)</td>
<td>903.90</td>
<td>1444.78</td>
<td>1448.34</td>
<td>476.18</td>
<td>434.81</td>
<td>1006.85</td>
<td>1742.75</td>
<td>-</td>
<td>0.5436</td>
<td>1.1193</td>
</tr>
<tr>
<td>Orkney Islands</td>
<td>3476.34</td>
<td>3693.42</td>
<td>3127.67</td>
<td>4679.14</td>
<td>4175.54</td>
<td>3739.02</td>
<td>2477.61</td>
<td>4206.50</td>
<td>-</td>
<td>1.4635</td>
</tr>
<tr>
<td>Irish</td>
<td>1652.35</td>
<td>2396.73</td>
<td>2124.06</td>
<td>1311.98</td>
<td>726.86</td>
<td>346.14</td>
<td>1897.29</td>
<td>1151.96</td>
<td>4016.56</td>
<td>-</td>
</tr>
</tbody>
</table>