

Northern glacial refugia for the pygmy shrew *Sorex minutus* in Europe revealed by phylogeographic analyses and species distribution modelling

Rodrigo Vega, Camilla Fløjgaard, Andrés Lira-Noriega, Yoshinori Nakazawa, Jens-Christian Svenning and Jeremy B. Searle

R. Vega (rvv9@cornell.edu) and J. B. Searle, Dept of Biology, Univ. of York, PO Box 373, York YO10 5YW, UK. (Present address of R. V.: Dept of Entomology, Comstock Hall 5123, Cornell Univ., Ithaca, NY 14853, USA.) – C. Fløjgaard, Ecoinformatics and Biodiversity Group, Dept of Biological Sciences, Aarhus Univ., Ny Munkegade 114, DK-8000 Aarhus C., Denmark and Dept of Wildlife Ecology and Biodiversity, National Environmental Research Inst., Aarhus Univ., Grenaavej 14, DK-8410 Ronde, Denmark. – A. Lira-Noriega and Y. Nakazawa, Natural History Museum and Biodiversity Research Center, The Univ. of Kansas, Lawrence, KS 66045, USA. – J.-C. Svenning, Ecoinformatics and Biodiversity Group, Dept of Biological Sciences, Aarhus Univ., Ny Munkegade 114, DK-8000 Aarhus C., Denmark.

The southern European peninsulas (Iberian, Italian and Balkan) are traditionally recognized as glacial refugia from where many species colonized central and northern Europe after the Last Glacial Maximum (LGM). However, evidence that some species had more northerly refugia is accumulating from phylogeographic, palaeontological and palynological studies, and more recently from species distribution modelling (SDM), but further studies are needed to test the idea of northern refugia in Europe. Here, we take a rarely implemented multidisciplinary approach to assess if the pygmy shrew *Sorex minutus*, a widespread Eurasian mammal species, had northern refugia during the LGM, and if these influenced its postglacial geographic distribution. First, we evaluated the phylogeographic and population expansion patterns using mtDNA sequence data from 123 pygmy shrews. Then, we used SDM to predict present and past (LGM) potential distributions using two different training data sets, two different algorithms (Maxent and GARP) and climate reconstructions for the LGM with two different general circulation models. An LGM distribution in the southern peninsulas was predicted by the SDM approaches, in line with the occurrence of lineages of *S. minutus* in these areas. The phylogeographic analyses also indicated a widespread and strictly northern-central European lineage, not derived from southern peninsulas, and with a postglacial population expansion signature. This was consistent with the SDM predictions of suitable LGM conditions for *S. minutus* occurring across central and eastern Europe, from unglaciated parts of the British Isles to much of the eastern European Plain. Hence, *S. minutus* likely persisted in parts of central and eastern Europe during the LGM, from where it colonized other northern areas during the late-glacial and postglacial periods. Our results provide new insights into the glacial and postglacial colonization history of the European mammal fauna, notably supporting glacial refugia further north than traditionally recognized.

During the Quaternary ice ages substantial areas of northern Europe were covered by ice sheets while permafrost existed in large areas of central Europe, which restricted the distribution of many temperate and warm-adapted species to the three southern European peninsulas of Iberia, Italy and the Balkans at the Last Glacial Maximum (LGM; Hewitt 2000). These species are interpreted to have recolonized central and northern Europe from these traditionally recognized southern glacial refugia in response to the late-glacial and postglacial warming (Taberlet et al. 1998, Hewitt 2000). Therefore, southern glacial refugia and the northward postglacial recolonization of central and northern Europe from these areas has become an established biogeographical paradigm (Hewitt 2000).

Other studies have, however, provided palaeontological, palynological and phylogeographic evidence that glacial refugia for some temperate and boreal species existed further north than the traditionally recognized southern European refugia, implying a more complex pattern of glacial survival and postglacial recolonization: fossils of temperate mammal species dated to the LGM (albeit rarely small mammals) have been described for a number of sites in central Europe, sometimes in co-occurrence with cold-adapted Pleistocene faunal elements (Sommer and Nadachowski 2006). Macrofossil charcoal (organic plant material ≥ 2 mm in diameter) of coniferous and broad-leaved trees dating to the Upper Palaeolithic has been found in several sites in Austria (42–23 Kya), Czech Republic (29–24.5 Kya), Croatia (27.8–10.8 Kya) and Hungary

(31.5–16.5 Kya), suggesting that these regions were also refugial areas for temperate deciduous species (Willis and van Andel 2004, Magri et al. 2006). Palynological records have shown European beech *Fagus sylvatica* pollen in several sites in central Europe between the late glacial and postglacial (15–10 Kya), and have shown that none of the three traditional refugial areas was the source for northern-central European beech populations (Magri et al. 2006). Phylogeographic studies on several small mammals have shown little similarity between Mediterranean and northern populations, and have described genetic clades linking together haplotypes sampled throughout northern-central Europe (Bilton et al. 1998, Kotlík et al. 2006). Furthermore, species distribution modelling (SDM) has shown that suitable climatic conditions existed for temperate and boreal species in northern latitudes supporting more northerly refugial areas in Europe (Svenning et al. 2008, Fløjgaard et al. 2009). However, a more comprehensive understanding of the relative importance of southern versus northern refugia in terms of LGM species' ranges as well as for postglacial recolonization is needed.

Here, we use the pygmy shrew *Sorex minutus* (Mammalia, Soricomorpha), as a model for studying the persistence of populations in northern European refugia during the LGM. *Sorex minutus* is widely distributed in the Palaearctic, throughout Europe to Lake Baikal (Siberia), including the three southern European peninsulas (Hutterer et al. 2008). The species occurs at low density in a wide range of terrestrial habitats with adequate ground cover (Churchfield and Searle 2008). In southern Europe the distribution becomes patchy and limited to higher altitudes 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.

Previous phylogeographic studies on *S. minutus* revealed a very widespread and genetically homogeneous “northern-central European and Siberian” lineage, extending from Britain through central and northern Europe to Siberia (ca 7000 km), but genetically distinct from the southern lineages in Iberia, Italy and the Balkans (Bilton et al. 1998, Mascheretti et al. 2003, McDevitt et al. 2010). These studies suggested that the northern-central European lineage persisted and expanded from one or more central or eastern European refugia located further north than the traditionally recognized southern European refugia. However, the size and locations of the possible northern refugia for *S. minutus* could not be assessed precisely.

Species distribution models combine information about species occurrences with environmental (usually climatic) data found across the study region to estimate the present-day geographical distribution of suitable environmental conditions for the species (Guisan and Zimmermann 2000). Then, the set of environmental conditions can be projected to past conditions to identify areas where there were suitable environmental conditions for the species (hindcasting) (Nogués-Bravo 2009), in this case at the LGM. Such SDM-based hindcasting has not been integrated into the previous phylogeographic studies on *S. minutus*, and the genetic data for central and eastern regions of Europe and in Siberia have been rather incomplete. This

makes it difficult to determine the importance of these regions for the LGM distribution of the species, its postglacial colonization history and its present-day genetic structure. Moreover, the inference of glacial refugia based solely on phylogeographic analyses can be obscured by the extinction of genetic variants, incomplete sampling and large-scale range shifts of the species (Waltari et al. 2007). Hence at this point, although the previous phylogeographic studies suggested the existence of northern glacial refugia for *S. minutus*, the size and geographic spread of these refugia as well as their role in the postglacial range dynamics of the species remain unclear.

The purpose of this study is to assess the distribution of *S. minutus* during the LGM based on a multidisciplinary approach using more detailed mtDNA-based phylogeographic analyses than conducted hitherto and including SDM-based hindcasting. Only a few studies have tried to estimate potential northern refugial areas in this way, despite the stronger inference allowed by these independent and highly complementary approaches (Waltari et al. 2007).

We assessed the following specific study questions: would a more detailed phylogeographic analysis also detect a distinctive “northern-central European and Siberian” lineage as has been previously found? Would this widespread lineage present a genetic signature of population expansion? Would different SDM-based hindcasting approaches predict suitable LGM conditions for *S. minutus* not only in the southern European peninsulas, but also further north, consistent with northern refugia? Would the combined phylogeographic and SDM approach allow us to estimate more precisely the geographic locations of northern refugia for *S. minutus*, as well as determine their potential role for its postglacial range dynamics? From the population expansion characteristics, how did the refugial populations colonize their current ranges? Finally, are the rather scant fossil data for *S. minutus* consistent with our phylogeographic and distributional findings?

This study sheds light on the spatial variation of the genetic diversity within the widespread distribution of *S. minutus*, its postglacial population expansion and colonization of Europe from northern refugia, and contributes towards an emerging new synthesis of the full-glacial distributions of the European biota. The nature of northern refugia also has important implications for the understanding of their biogeographic roles as sources of genetic diversity, areas of speciation, identification of conservation units and preservation of species, particularly in response to future climate change (Kotlík et al. 2006, Provan and Bennett 2008).

Materials and methods

Phylogeographic analyses

Samples and laboratory procedures

In total, 123 individuals of *S. minutus* from Europe and Siberia were used for the phylogeographic analysis of the mitochondrial cytochrome b (cyt b) gene. Sixty-six *S. minutus* cyt b sequences were obtained from Genbank

(AB175132; Ohdachi et al. 2006; AJ535393–AJ535457; Mascheretti et al. 2003). Fifty-seven out of the 123 samples of *S. minutus* were obtained from northern-central Europe during fieldwork and from museum collections (see Acknowledgements) to increase the molecular data and to provide a more detailed analysis of this region. A sequence of *S. volnuchini* was used as outgroup (AJ535458; Mascheretti et al. 2003).

Genomic DNA was extracted using a commercial kit (Qiagen). Partial cyt b sequences were obtained by PCR using two primer pairs that amplified ca 700 bp of overlapping fragments. 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 Univ. Core Laboratories Center).

Sequence and phylogenetic analyses

Cyt b sequences were edited in BioEdit 7.0 (Hall 1999) and aligned by eye. For the construction of phylogenetic trees, the model of evolution that best fitted the molecular data was searched using MrModeltest 2.3 (Nylander 2004) using the minimum Akaike information criteria value. The substitution model supported was the General Time Reversible with specified substitution types (A–C = 0.3663, A–G = 17.4110, A–T = 1.0216, C–G = 2.1621, C–T = 13.0604, G–T = 1.0), proportion of invariable sites (0.5332), gamma shape parameter (0.9799) and nucleotide frequencies (A = 0.2750, C = 0.2996, G = 0.1382, T = 0.2872).

The phylogenetic relationships within *S. minutus* were inferred by Neighbour-Joining (NJ), Maximum Likelihood (ML) and Bayesian analysis using PAUP* 4.0b10 (Swofford 2000), PhyML 3.0 (Guindon and Gascuel 2003) and MrBayes 3.1 (Huelsenbeck and Ronquist 2001), respectively. Confidence for the phylogenetic relationships in NJ and ML was assessed by bootstrap replicates (10 000 and 500 replicates, respectively). For the Bayesian analysis, two independent runs were performed with 10 million generations and 5 chains each, a sampling frequency every 1000 generations, a temperature of 0.1 for the heated chain and checking for convergence. Trees were summarized after a burn-in value of 2500 to obtain the posterior probabilities of each phylogenetic branch.

Phylogenetic networks provide an explicit graphic representation of evolutionary history between sequences in which taxa are represented as nodes and their evolutionary relationships are represented by edges. Most internal nodes represent ancestral states from which more recent and peripheral nodes derive (Avice 2000). A parsimony phylogenetic network of cyt b haplotypes was constructed using the software Network 4.5 (Fluxus-Engineering) with a median-joining algorithm and a greedy FHP genetic distance calculation method. The median joining algorithm identifies groups of haplotypes and introduces hypothetical (non-observed) haplotypes to construct the parsimony network.

Genetic and statistical analyses

Standard sequence polymorphism indices (number of haplotypes, polymorphic sites and parsimony informative sites) and genetic diversity values (π , nucleotide diversity \pm SD; h , haplotype diversity) were estimated using Arlequin 3.11 (Excoffier et al. 2005).

Population expansion was examined for both the full dataset (Eurasia) and for the “northern-central European and Siberian” lineage using DnaSP 5.0 (Librado and Rozas 2009). In each case a mismatch distribution (distribution of the number of differences between pairs of haplotypes) was estimated to compare the demography of the populations with the expectations of a sudden population expansion model (Rogers and Harpending 1992). The raggedness index (rg), which measures the smoothness of the observed distribution, was computed and the statistical validity of the estimated expansion model was tested by a parametric bootstrap approach as a sum of square deviations (SSD) between the observed and the expected mismatch (Schneider and Excoffier 1999) using Arlequin (10 000 replicates). Three other tests for population expansion were performed in DnaSP using coalescent simulations to test for statistical significance (10 000 replicates): R_2 test of neutrality, based on the difference of the number of singleton mutations and the average number of nucleotide differences (Ramos-Onsins and Rozas 2002); Fu's F_s , a statistic based on the infinite-site model without recombination that shows large negative F_s values when there has been a demographic population expansion (Fu 1997); Tajima's D , a test for selective neutrality based on the infinite-site model without recombination where significant values appear from selective effects but also from factors such as population expansion, bottleneck or heterogeneous mutation rates (Tajima 1989).

Species distribution modelling

Important discrepancies in the prediction of the potential distribution of a particular species arise from differences in data sample size (Stockwell and Peterson 2002, Wisz et al. 2008), environmental and/or climatic data (Peterson and Nakazawa 2008), and algorithms (Peterson et al. 2007, but see Phillips 2008). Also, if the occurrence records used to model the distribution do not adequately sample the environmental requirements of the species, the prediction will not truly reflect its potential geographic distribution (Pearson et al. 2007). Therefore, to ensure the robustness of our findings, we modelled the potential distribution of *S. minutus* in the present and at the LGM using two independent training data sets, two algorithms, namely the maximum entropy algorithm (Maxent; Phillips et al. 2006) and the Genetic Algorithm for Rule-set Prediction (GARP; Stockwell and Noble 1992, Stockwell 1999), and using climate reconstructions for the LGM based on two general circulation models (GCMs). All GIS operations were performed using ArcGIS 9.3 (ESRI, Redlands, CA, USA).

Species occurrence data

For the first data set, hereafter termed “data set 1”, we used the species records from fieldwork, from two online sources (Global Biodiversity Information Facility, GBIF,

and Mammal Networked Information System, MaNIS) and from museum specimens obtained for our study (see Acknowledgements). Most of the data were derived from the following sources: the Atlas of Mammals in Britain (Arnold 1993), the European Environment Agency, the UK National Biodiversity Network, the Highland Biological Recording Group – HBRG Mammals data set and the Ministerio de Medio Ambiente y Medio Rural y Marino (Spain). Low precision occurrences, such as presence data taken from the centroids of atlas grids and falsely georeferenced occurrences (i.e. offshore and out-of-range locations), were eliminated from this data set. In total, we collected 536 high-precision unique latitude-longitude localities, but this data set was geographically biased towards western Europe and Britain due to differences in sampling effort across the species' distribution range (i.e. there are few species records from Siberia and southern Europe). In order to correct for sampling bias, we created 25 random subsets from the original data set to limit the number of unique occurrences to ≤ 5 in squares of 5×5 degrees distributed across the extent of the geographical analysis (Wisz et al. 2008). This procedure yielded a total of 146 unique localities for each subset which were more evenly distributed.

For the second data set, hereafter termed “data set 2”, we used the records from the Atlas of European Mammals (AEM; Mitchell-Jones et al. 1999) which present less geographic bias within Europe, but had a much coarser resolution than data set 1. The AEM uses an approximate equal area grid of 50×50 km based on the Universal Transverse Mercator (UTM) projection and the Military Grid Reference System (MGRS). Records of “species presence” as well as “presence assumed” (i.e. presence was observed before 1970 and no evidence of later extinction) were included in the study and a total of 1178 data points were used.

To ensure transferability of our models, we used a geographically independent test data set. We digitized the Eurasian range map for *S. minutus* (Hutterer et al. 2008) and recorded the species as present in all 50×50 km MGRS grid cells within the outline of the range map. Then, we used the part of the range located east of the European study area (for simplicity referred to hereafter as Siberia) only as a test data set ($n = 3122$ data points). This allowed us to evaluate the performance of the models with both data sets and assess which climatic variables provided the strongest predictive ability in a geographically independent region with relatively LGM-like conditions (Fløjgaard et al. 2009). We used the digitised range map data only for testing, given its much coarser resolution and uncertain quality compared to the occurrence data from data sets 1 and 2.

Climate data

For the present-day SDM we initially considered the 19 bioclimatic variables from the WorldClim dataset at a spatial resolution of 2.5 minutes <www.worldclim.org/>. These climate layers are based on spatially interpolated values of temperature and precipitation gathered from weather stations around the world from 1950–2000 (Hijmans et al. 2005). For the LGM (21 Kya) we used

the climate reconstructions of the same 19 bioclimatic variables based on the CCSM3 (Collins et al. 2006) and MIROC3.2 (Hasumi and Emori 2004) GCMs <<http://biogeo.berkeley.edu/>> at a spatial resolution of 2.5 minutes.

We used the Jackknife procedure implemented in Maxent with the 19 bioclimatic variables on the two data sets to find the best set of predictor variables. We assessed the performance of the models based on the Area Under the Curve (AUC) values of the Receiver Operating Characteristic (ROC) in the independent test region of Siberia. The worst predictor of the whole set of variables was eliminated, a new model was produced using the remaining variables and the process was repeated until all variables were exhausted. We chose the final set of predictors based on parsimony (i.e. with the fewest number of climatic variables) and with the highest AUC value in the independent test region of Siberia.

The final set of predictors comprised the variables Annual Mean Temperature (AMT) and Precipitation of the Warmest Quarter (PWQ); thus, AMT and PWQ were used for estimating the present and LGM distribution of *S. minutus*. These two variables were not highly correlated ($r = -0.3550$) and models that included only these yielded higher or almost equal AUC values than models that included only one or more variables in combination with AMT and PWQ. In addition, these variables are biologically meaningful for *S. minutus* considering its broad distribution in northern-central Europe and Siberia and habitat preference for damp and temperate areas (Churchfield and Searle 2008, Hutterer et al. 2008). The modelling was performed with data sets 1 and 2 as inputs in Maxent and GARP, and all models were evaluated on the geographically independent (extrinsic) test data from Siberia. For data set 1 we made models with all 25 subsets. Finally, all models were projected onto the two LGM climate reconstructions to identify the potential distribution of *S. minutus*.

Modelling algorithms

To assess the variation in the outcome of model predictions due to differences in modelling algorithms, we used Maxent and GARP. Maxent has been shown to perform very well in comparative studies of species distribution modelling compared to GARP (Elith et al. 2006, Phillips and Dudík 2008, Elith and Graham 2009, but see also Peterson et al. 2008), while GARP has been shown to perform better than Maxent in transferability studies (Peterson et al. 2007, but see also Phillips 2008). Ultimately, the performance of each algorithm may be properly compared using the corresponding thresholding during model evaluation, since their predictions are not given in the same scale (Peterson et al. 2008).

To evaluate the accuracy of our models, the empirical AUC values were compared against the AUC values of 1000 random models, as implemented in Peterson et al. (2008), using the data from the test region. AUC ROC values are expressed as the ratio of the area under the observed curve (i.e. the overall area for which each algorithm predicts as present) to the area under the line that defines a random expectation; consequently, the AUC values are expected to

be larger than one as the model departs from the random expectation (Peterson et al. 2008).

Maxent is a machine-learning technique based on the principle of maximum entropy that fits a probability distribution to the environmental conditions at the locations where a species has been observed (Phillips et al. 2004, 2006). When implemented with ecologically meaningful sets of predictor variables, Maxent produced similar estimates for the locations of glacial refugia as Bioclim, another commonly used, but simpler, modelling technique (Svenning et al. 2008, Fløjgaard et al. 2009). We used the default settings in Maxent 3.2.1 <www.cs.princeton.edu/~schapire/maxent/> with background data limited to Eurasia as described in the species occurrence data section. We converted the continuous logistic output from Maxent into a binary map of predicted suitable environmental conditions for *S. minutus* using the maximum test sensitivity and specificity threshold because it optimized the correct discrimination of presences and pseudoabsences in the test data.

GARP is a genetic algorithm that produces a set of rules that describe the non-random association between environmental variables and occurrence data (Stockwell and Noble 1992, Stockwell 1999). First, the algorithm creates a set of rules based on four basic types (bioclimatic, atomic, negated and logistic regression rules), their individual predictive accuracy is calculated and only those rules with the highest predictive accuracy are retained in the model. The overall performance of the model is evaluated using a subset of presence points. Then, a second generation of rules is produced via the random modification of the previous generation rules, their predictive accuracy is calculated and only those with the highest accuracy are included in the model. Finally, the overall performance of the model is re-evaluated and the process of creation, evaluation and inclusion of rules is repeated until a maximum number of iterations is reached (1000 in this case), or until performance values no longer change appreciably from one iteration to the next (convergence parameter of 1%). We used the version of DesktopGarp as implemented in openModeller ver. 1.0.9 <<http://openmodeller.sf.net>> using the default parameters (Anderson et al. 2003). We converted the continuous output into a binary map of predicted occurrence of the suitable conditions for *S. minutus* by assigning a value of 1 for the model values that corresponded to 10% or more of the testing points.

Results

Phylogeographic analysis

Sequence analysis and phylogenetic reconstructions

A partial sequence of 1110 bp from the *S. minutus* cyt b was analysed. One hundred and twelve haplotypes were obtained, from which 46 were newly described and deposited in GenBank (accession numbers: GQ494305–GQ494350). There were 894 invariable and 216 variable positions, from which 137 were parsimony informative.

All the phylogenetic analyses revealed five distinct lineages (Fig. 1). Samples from the Mediterranean peninsulas clustered in three lineages, namely the “Iberian”, “Italian” and “Balkan” groups, corresponding to their

geographical origin. There was also a well supported “Pyrenean” lineage with samples from Andorra and Ireland. Samples from northern-central Europe and Siberia clustered together forming a geographically widespread lineage that did not include any individuals from the southern peninsulas, hereafter named as the “northern-central European” lineage. This lineage was composed of 105 sequences (94 haplotypes) with 940 invariable and 170 variable positions, from which 92 were parsimony informative.

The phylogenetic network of the northern-central European lineage presented a star-like pattern with three most central haplotypes, named A, B and C, separated by only one mutational step from each other and from which all other sequences derived (Fig. 2). The other phylogroups from the southern peninsulas were much more distantly related and separated by several mutations (data not shown). The central haplotypes A and B were entirely composed of samples from the Netherlands (three and two individuals, respectively), while the third central haplotype (C) belonged to a central Ukrainian specimen from the locality Tishki (50°6.27'N, 33°6.39'E). There was an apparent geographical subdivision of the samples that were connected to these three central haplotypes (Fig. 2). Only haplotypes from Great Britain and the Netherlands were directly connected to A. Several haplotypes from different countries of northern and central Europe were connected to B, also including some haplotypes from Great Britain and the Netherlands, but there were no haplotypes from eastern Europe or Siberia (except for one sample from Ukraine ambiguously connected to B and C). Haplotypes from northern, central and eastern Europe and Siberia were all directly connected to C, but there were no samples from countries further west than Germany. However, the support for these subdivisions was not strong: equally parsimonious explanations (loops) appeared in the central part of the network between B and C, and there was no supported sub-structure within the northern-central European lineage in the phylogenetic trees.

Genetic and statistical analyses

The whole Eurasian sample presented a nucleotide diversity $\pi = 0.0109 \pm 0.0055$, and a haplotype diversity $h = 0.9983$. The northern-central European lineage had a nucleotide diversity $\pi = 0.0067 \pm 0.0035$, and a haplotype diversity $h = 0.9980$. Genetic diversity values were not calculated for the southern European lineages because of small sample size.

The mismatch distribution of the whole dataset (Eurasia) was bimodal, consistent with pairwise differences between sequences belonging to the same and different lineages (Fig. 3a). The mismatch distribution of the northern-central European lineage showed a unimodal distribution that, visually, fitted almost perfectly over the expected values for a population expansion model (Fig. 3b). There was an observed mean of 7.382 pairwise differences with a variance of 8.152. The goodness of fit test showed no significant differences between the observed and expected values under a sudden expansion model for the northern-central European lineage ($SSD = 0.0004$, $p_{SSD} > 0.05$; $rg = 0.0082$, $p > 0.05$). Negative and significant Tajima's D ($D = -2.5721$, $p < 0.001$) and Fu's F_s ($F_s = -24.8437$,

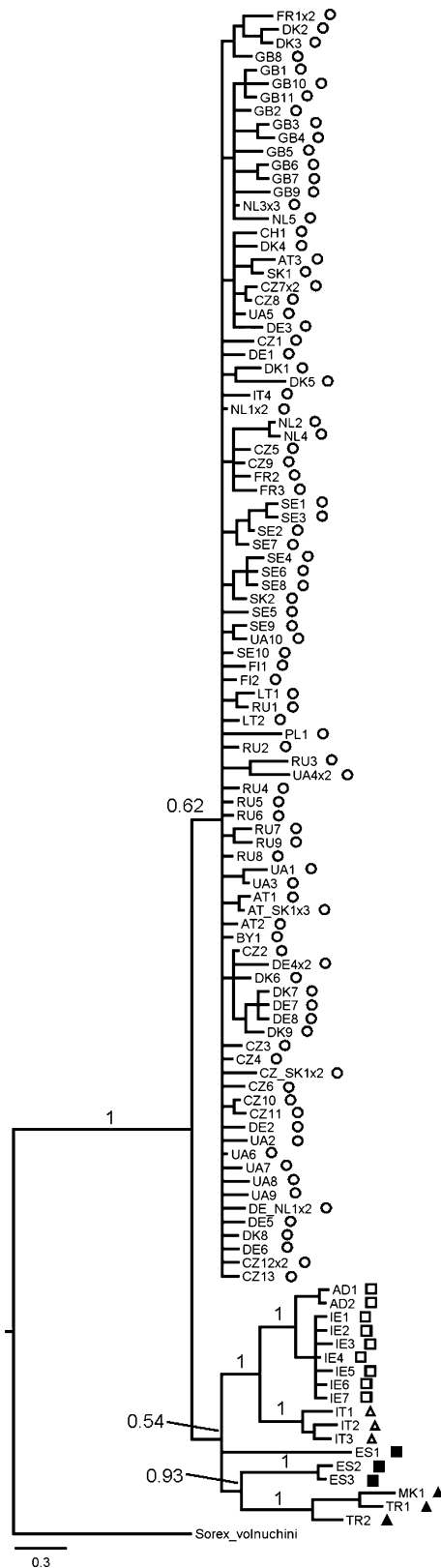


Figure 1. Continued.

$p < 0.001$) showed departures from neutrality also consistent with a sudden population expansion. Moreover, the R_2 test of neutrality also showed that the northern-central European lineage gave a genetic signature consistent with a

sudden population expansion ($R_2 = 0.0180$, $p < 0.001$). The rest of the sequences and lineages that belonged to the more distantly related southern European lineages (Iberian, Italian and Balkan peninsulas) and the Pyrenees were not analysed because of small sample size.

Species distribution modelling

Predicted present distribution

Species distribution models from Maxent matched the reported distribution of the species (Fig. 4a, c). The models also predicted suitable climatic conditions outside the reported distribution of the species especially in two regions, the Asia Minor-Caucasus region and in the Far East (Fig. 4a, c). The predicted present distribution of *S. minutus* with GARP was very similar to that of Maxent, it also matched the reported distribution and the predicted suitable climatic conditions in the Asia Minor-Caucasus region and in the Far East (Fig. 4b, d).

All Maxent and GARP models were accurate in the test region, with AUC values for both data sets higher than null expectations ($p < 0.001$; mean $AUC_{MAXENT} = 1.24 \pm 0.021$ and mean $AUC_{GARP} = 1.049 \pm 0.007$ for data set 1, and mean $AUC_{MAXENT} = 1.249 \pm 0.011$ and mean $AUC_{GARP} = 1.032 \pm 0.005$ for data set 2).

Predicted LGM distribution

With the two data sets and GCMs, Maxent and GARP predicted suitable LGM climatic conditions in the southern European peninsulas (Fig. 4e–l), concordant with southern refugia. In general, suitable LGM conditions with the two data sets, GCMs and algorithms were also predicted north of the southern refugia, particularly throughout central Europe, most of eastern Europe, southern Poland, eastern and southern Ukraine, the Crimea peninsula and the Caucasus. With Maxent, the LGM predictions differed little between data sets or between GCMs, and there were predicted suitable conditions in central and eastern Europe close to the ice sheet (Fig. 4e, g, i, k). With GARP, predictions differed between GCMs: more restricted suitable conditions in central and eastern Europe were predicted with CCSM3 (Fig. 4f, h) than with MIROC3.2 (Fig. 4j, l), but predictions did not differ much between data sets. The most restricted predictions (using GARP with CCSM3) still showed suitable climatic conditions in southern Ireland, central and southern France, western parts

Figure 1. Bayesian inference tree showing the phylogenetic relationships among *Sorex minutus* samples (*S. volnuchini*, out-group). Five lineages were found (\square = Pyrenean-Irish, Δ = Italian, \blacksquare = Iberian, \blacktriangle = Balkan, and \circ = northern-central European). The northern-central European lineage is geographically widespread but has not been found within the southern European peninsulas. Values on branches correspond to Bayesian posterior probabilities. Haplotypes are represented with two-letter country codes followed by an identification number (x2, haplotype frequency = 2 etc.): AD = Andorra, AT = Austria, BY = Belarus, CH = Switzerland, CZ = Czech Republic, DE = Germany, DK = Denmark, ES = Spain, FI = Finland, FR = France, GB = Great Britain, IE = Ireland, IT = Italy, LT = Lithuania, MK = Macedonia, NL = the Netherlands, PL = Poland, RU = Russia, SE = Sweden, SK = Slovakia, TR = Turkey, UA = Ukraine.

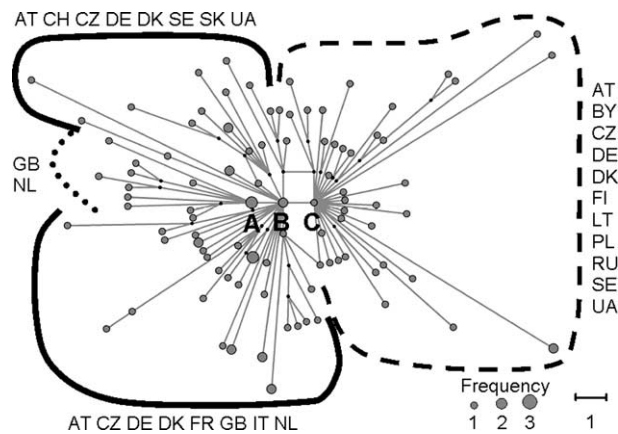


Figure 2. Parsimony median joining haplotype network for the northern-central European lineage of *Sorex minutus*. Observed haplotypes are shown as grey circles (proportional to frequency) and hypothetical haplotypes are shown as black circles. There is a star-like phylogeny with three central (ancestral) haplotypes. A and B are two central haplotypes from the Netherlands, and C is from central Ukraine. The dotted black line encircles haplotypes directly linked to A, black lines encircle haplotypes directly linked to B and the dashed line encircles haplotypes directly linked to C (the country of origin for haplotypes is shown next to clusters; two-letters country codes as in Fig. 1). For simplicity, haplotypes from the more diverged southern European lineages are not shown, but relate to central-European haplotypes by the addition of several hypothetical haplotypes and >10 mutational steps. The scale bar represents one mutational step.

of Switzerland, a few regions north of the Balkans, the Crimea peninsula and the Caucasus.

Discussion

Northern glacial refugia revealed by a combined approach

Sorex minutus is considered a temperate species, but it is also latitudinally distributed above 60°N (i.e. near the Arctic Circle) and altitudinally above 2000 m in regions with permafrost and harsh winters (Mitchell-Jones et al. 1999, Hutterer et al. 2008). Northern non-arctic species like *S. minutus* could have persisted in high latitude refugia in Europe during the LGM, north of the traditionally recognized Mediterranean refugial areas (Stewart and Lister 2001). This could have been a result of their ecological traits (notably cold tolerance) and biogeographical characteristics that may have determined their response to the glaciations (Bhagwat and Willis 2008). *Sorex minutus* is, therefore, a suitable model organism for exploring the controversial hypothesis of “northern” glacial refugia.

The general concordance of the phylogeographic analyses with the predicted LGM distributions based on species distribution modelling and the concordance between models suggest that we have obtained robust results concerning the LGM distribution of *S. minutus*. Our phylogeographic analyses provided evidence for a distinct lineage in northern-central Europe, with additional lineages in the Iberian, Italian and Balkan peninsulas in southern Europe. First, the absence of southern haplotypes in northern-central Europe supports the hypothesis that the

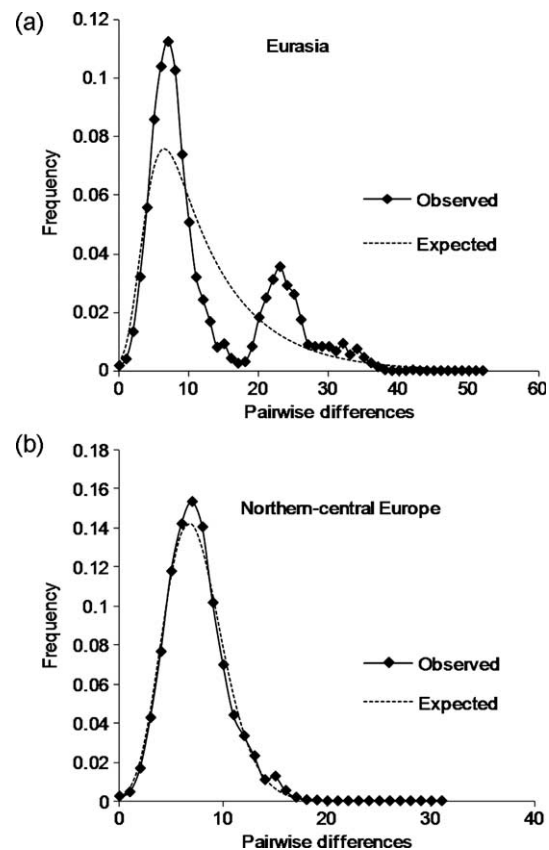


Figure 3. Mismatch distribution for observed (continuous line) and expected (dashed line) pairwise comparisons under a sudden population expansion model among *Sorex minutus* cyt b sequences. (a) Mismatch distribution among Eurasian sequences with a bimodal observed distribution where the first peak corresponds to pairwise comparisons among closely related individuals within lineages, while the second peak corresponds to pairwise comparisons among distantly related individuals from different lineages. (b) Mismatch distribution among sequences from the northern-central European lineage showing a unimodal distribution, a genetic signature which corresponds to the expected distribution for sudden population expansion.

southern peninsulas were areas of endemism and differentiation for *S. minutus*, but not for northward colonization (Bilton et al. 1998), i.e. the current populations in northern-central Europe were not derived from LGM populations in the traditional southern European refugia. Second, the northern-central European lineage showed a strong signature of population expansion supported by the mismatch distribution and population expansion tests. Finally, ancestral haplotypes in a phylogenetic network can be identified by their central or internal position from where the peripheral, more recent, haplotypes are derived, by the number of haplotypes that arise from them and by their abundance (Avice 2000). The phylogenetic network of the northern-central European lineage showed a star-like pattern with three ancestral haplotypes from distant regions in central and eastern Europe (the Netherlands and Ukraine). This pattern was also consistent with a widespread LGM distribution and congruent with the hypothesis of persistence and postglacial expansion from northern glacial refugia.

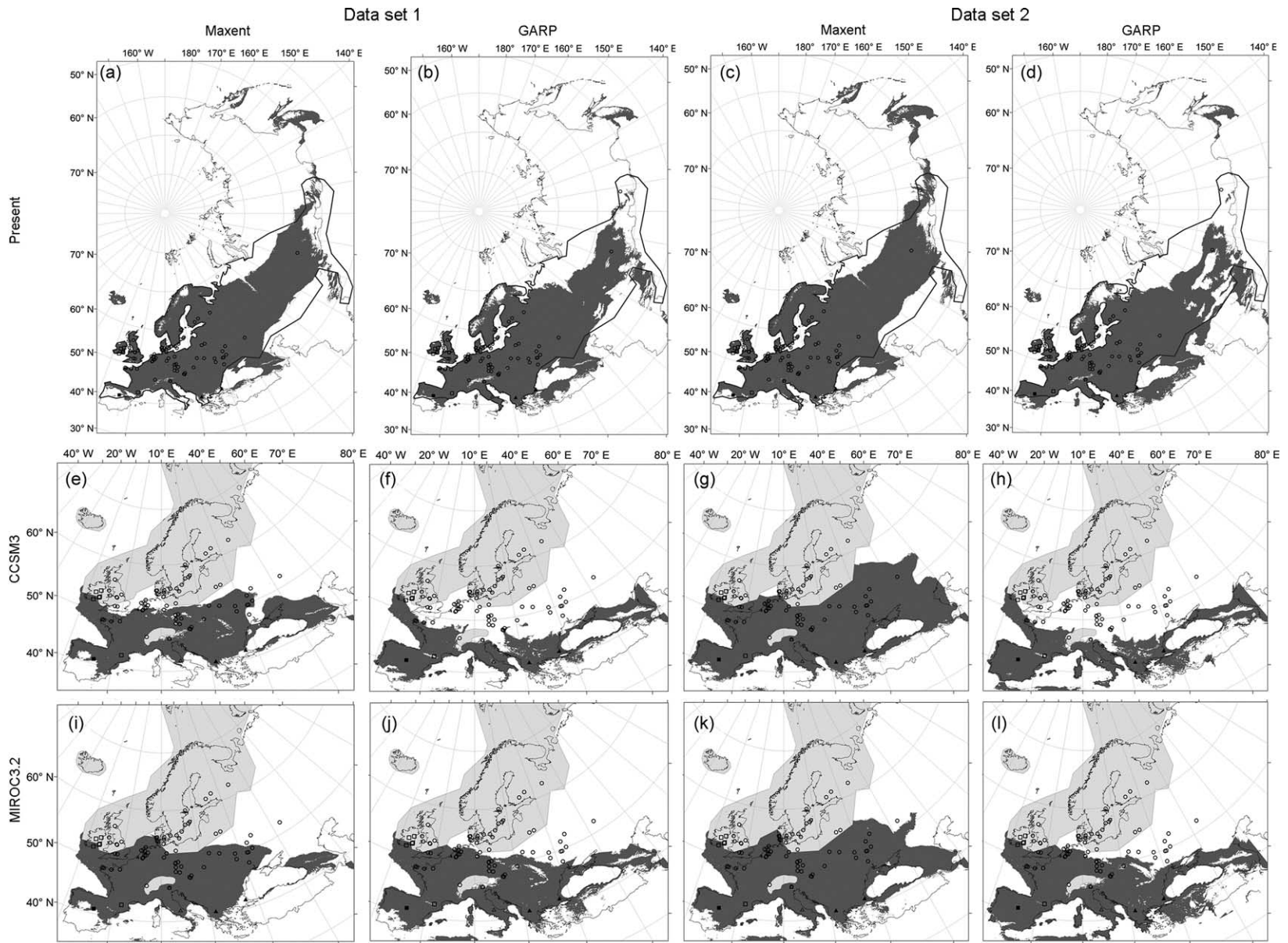


Figure 4. Continued.

The phylogeographic pattern that we observe here did not arise from the low sample size in southern Europe: the few samples from southern peninsulas belonged to lineages differentiated by a large number of mutation steps from the northern-central European lineage; if northern-central Europe had been colonized from southern Europe we would have found northern-central European samples clustering within southern lineages, not forming a separate lineage. Moreover, a phylogeographic study on *S. minutus* using the mitochondrial Control Region and Y-chromosome introns with more samples from southern peninsulas showed a similar pattern (McDevitt et al. 2010). Nevertheless, further sampling in southern regions and the use of other molecular markers is desirable to investigate the genetic variation and population expansion events within Mediterranean peninsulas, and for the determination of contact zones among lineages.

We did not use the mismatch distributions to date the population expansion for the northern-central European lineage because of the lack of a suitable mutation rate for *cyt b* in *S. minutus*. Previous studies on *Sorex* have used mismatch distributions for molecular dating (e.g. Ratkiewicz et al. 2002), but with mutation rates that may not be suitable over short time frames (Ho et al. 2005).

The modelling approaches predicted successfully the wide present-day distribution of *S. minutus* in Eurasia. Therefore, we consider our SDM approaches as giving realistic estimates of the area with suitable climatic conditions for our species and of its potential LGM distribution. A third model using Bioclim with SDM data sets 1 and 2 also resulted in very similar present-day and LGM distributions for *S. minutus* (data not shown). The potential LGM distributions predicted by our SDM approaches not only included the traditionally recognized southern refugia, but also a wide area across central and eastern Europe, from the unglaciated parts of southern Ireland and Britain to most of the central and southeast European (or Russian) Plain. In particular, the predicted LGM distribution throughout central and eastern Europe encompasses suggested northern refugial areas based on palaeontological and palynological data for other temperate and boreal species (Willis et al. 2000, Willis and van Andel 2004, Magri et al. 2006, Sommer and Nadachowski 2006). Thus, the northern-central European lineage could have persisted in various parts of this wide area during the LGM according to the phylogeographic and the SDM approaches.

We note that the central and eastern European LGM distribution was similar with both data sets, particularly when using Maxent (with both GCMs) and when using GARP with MIROC3.2, even though we used very different species records. However, the LGM distributions

when using GARP were more widespread to the north with MIROC3.2 than with CCSM3 GCMs, which could represent variations due to modelling algorithms and GCMs. Also, the predicted present-day suitable climatic conditions outside the reported distribution of *S. minutus* in the Asia Minor-Caucasus region and in the Far East probably reflect competitive or speciation processes rather than an inaccurate estimation of the suitable climatic conditions. In Asia Minor-Caucasus, *S. minutus* is replaced by the closely related sister species *S. volnuchini*, while in the Far East many other *Sorex* species occur including similar-sized species such as *S. gracillimus*.

The predicted LGM distribution of *S. minutus* appears to be continuous throughout Europe; however, lineage diversification is still plausible: First, the present distribution of *S. minutus* also appears to be continuous but it is affected by landscape features, not evident at the geographic resolution given, which could have subdivided the species range. Therefore, it could be expected that landscape features at the LGM also affected the distribution of *S. minutus*. Second, the estimation of the extent of ice sheets in mountainous areas is not precise, so it may be expected that the Iberian and Italian populations remained isolated from the rest of Europe by ice sheets covering the Pyrenees and the Alps, respectively, while the heterogeneous landscape in the Balkans could have been responsible for the limited distribution of the genetic lineage there. Also, different genetic variants could have arisen within regions and could have been maintained there selectively reducing further spread into contiguous regions. Another explanation could be that interspecific competition and/or other non-climatic conditions subdivided the potentially continuous LGM distribution.

Insights into postglacial colonization

The predicted distribution for *S. minutus* in the Iberian, Italian and Balkan peninsulas presumably corresponds to the refugial areas where the southern genetic lineages persisted during the LGM. The Pyrenean lineage, here represented by a limited number of Andorran and Irish samples, could have persisted during the LGM in central and south-western France and even in unglaciated areas in southern Ireland, as shown by our SDM models. However, genetic studies support a more recent origin of the Irish pygmy shrew, transported there by humans during the Holocene (Mascheretti et al. 2003, McDevitt et al. 2009, A. D. McDevitt, V. R. Rambau, R. Vega and J. B. Searle pers. comm.). Further molecular sampling in southern Europe is desirable to determine the extent of the

Figure 4. Species distribution modelling of *Sorex minutus* in the present and at the Last Glacial Maximum (LGM) using different approaches. Two independent data sets, two algorithms, Maximum entropy (Maxent) and Genetic Algorithm for Rule-set prediction (GARP), and climate reconstructions for the LGM based on two general circulation models (CCSM3 and MIROC3.2) were used. Climatic variables were obtained from WorldClim and two were selected as best predictors with a Jackknife procedure: annual mean temperature and precipitation of the warmest quarter. (a–d) Maxent and GARP modelled present distributions with data sets 1 and 2. (e–l) Maxent and GARP modelled LGM distributions with data sets 1 and 2 using CCSM3 and MIROC3.2. The thick lines (a–d) represent the outline of present-day distribution range of the species, the dark shading corresponds to present-day and LGM suitable climatic conditions, and the light gray polygon represents the ice extent at the LGM, about 21 Kya (redrawn from Svendsen et al. 2004). Location of samples used for the phylogeographic analysis is shown (lineages as in Fig. 1: □ = Pyrenean-Irish, △ = Italian, ■ = Iberian, ▲ = Balkan, and ○ = northern-central European).

geographic distribution of the lineages found there and the contact zones between them.

Considering the phylogenetic network for the northern-central European lineage, the three central (ancestral) haplotypes were located in or near regions where the SDM approaches predicted a potential LGM distribution for *S. minutus*. These results imply that *S. minutus* was not dependent on amelioration of the climate at the end of the last glaciation to colonize northern-central Europe from southern refugia; instead, it was already present. As the ice sheets retreated and the climate improved, the range of *S. minutus* expanded from northern refugia colonizing the rest of northern-central Europe. For example, Scandinavian and the Baltic regions were most likely colonized by pygmy shrews from eastern Europe, not from the west or from southern peninsulas. Thus, the phylogenetic network shows that sequences from Norway, Finland and Lithuania group closely with the Ukrainian central haplotype, which according to the SDM modelling could have survived the LGM in situ on the east European Plain. Likewise, the genetic similarity of samples from the Netherlands and Britain, in comparison to those elsewhere, suggests that the British pygmy shrew originated from populations in the vicinity of the Netherlands, reaching Britain over the landbridge with continental Europe. An alternative explanation is that *S. minutus* persisted in the unglaciated regions of southern Britain (as predicted by several of our SDM approaches) which were geographically connected and genetically similar to populations in continental Europe during the LGM. Whatever the explanation, as ice sheets retreated, *S. minutus* belonging to the northern-central European lineage was able to colonize the northern parts of mainland Britain.

Further support from fossils and phylogeographic analyses

Northern refugia in central Europe and further east, north of the traditional Mediterranean refugia, have been hypothesized in phylogeographic analyses for a number of small mammal species other than the pygmy shrew, including the field vole *Microtus agrestis* (Jaarola and Searle 2002), bank vole *Clethrionomys glareolus* (Deffontaine et al. 2005, Kotlík et al. 2006), root vole *Microtus oeconomus* (Brunhoff et al. 2003), common vole *Microtus arvalis* (Heckel et al. 2005) and the common shrew *Sorex araneus* (Bilton et al. 1998, Yannic et al. 2008). For bank voles, root voles, field voles and common voles, predictions of their potential LGM distribution based on SDM were also consistent with northern refugia (Fløjgaard et al. 2009).

Most of the phylogeographic studies point to the Carpathians as a likely northern refugial area, but a refugium in this area could have included broader regions of Hungary, Slovakia, Czech Republic, Moldova and Poland, supported by the occurrence of temperate mammal fossil records in the area (Sommer and Nadachowski 2006) and by our results. Also, 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 refugium north of

the traditionally recognized southern refugia (Sommer and Nadachowski 2006), further supported by our findings.

In addition, there are a few but important fossil records of *S. minutus* from several localities north of the southern refugia, radiocarbon dated close to the LGM or earlier (S3P Faunal Database <www.esc.cam.ac.uk/research/research-groups/oistage3/stage-three-project-database-downloads>). These fossil remains have been found in sites in France (26 Kya), Belgium (38–40 Kya), Germany (23–29 Kya) and Hungary (20–22 Kya).

In conclusion, a wide northern LGM distribution for *S. minutus* is supported by the combined use of a phylogeographic and species distribution modelling approach. The SDM methodologies provide evidence for a central and eastern European LGM distribution of *S. minutus*, where the northern-central European lineage could have been distributed. Additionally, the SDM approaches reveal potential LGM distributions for *S. minutus* in southern refugia, consistent with the lineages present in those regions. The phylogeographic analyses, however, indicate that the southern refugia were not the postglacial source of the current and widespread northern-central European populations. The other phylogeographic and SDM studies on small mammals, mammal and plant fossil records, and *S. minutus* fossil remains presented here provide additional evidence consistent with or directly supportive of our findings.

Our results contribute to the understanding of persistence and colonization from glacial refugia further north than traditionally recognized. They also provide new insights into the location and importance of refugial areas for the persistence of populations and genetic lineages during climate change. The use of *S. minutus* as a model exemplifies how the combined use of phylogeography and species distribution modelling can be applied to understand present-day biodiversity patterns, and can predict and test the past distribution of species to gain insight into the colonization patterns, differentiation and biogeography of species.

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: Dept of Ecology and Evolution (Univ. de Lausanne, Switzerland), Univ. na Primorskem and Research Centre of Koper (Slovenia), Natuurhistorisch Museum (Rotterdam, the Netherlands), Dipartimento di Ecologia (Univ. della Calabria, Italy), Museo di Anatomia Comparata and Museo di Zoologia “La Sapienza” (Univ. di Roma, Italy) and Natuurmuseum Brabant (Tilburg, the Netherlands). We are very grateful for the tissue samples provided by Boris Kryštufec, Allan McDevitt, Glenn Yannic, Jacques Hausser, Jan Zima, Friða Jóhannesdóttir, Holger Bruns, Peter Borkenhagen and Petr Kotlík. We thank David Nogués-Bravo and two anonymous referees for their valuable comments. Bayesian analyses were run at the Computational Biology Service Unit from Cornell Univ. which is partially funded by Microsoft Corporation. We gratefully acknowledge financial support to R. Vega (181844) and A. Lira-Noriega (189216) from CONACyT (México), and to J.-C. Svenning from the Danish Natural Science Research Council (grant 272-07-0242). This work represents the fruits of the discussion of our work presented at the 4th Biennial

References

- Anderson, R. P. et al. 2003. Evaluating predictive models of species' distributions: criteria for selecting optimal models. – *Ecol. Model.* 162: 211–232.
- Arnold, H. R. 1993. Atlas of mammals of Britain. – Her Majesty's Stationary Office, London.
- Avise, J. C. 2000. Phylogeography: the history and formation of species. – Harvard Univ. Press.
- Bhagwat, S. A. and Willis, K. J. 2008. Species persistence in northerly glacial refugia of Europe: a matter of chance or biogeographical traits? – *J. Biogeogr.* 35: 464–482.
- Bilton, D. T. et al. 1998. Mediterranean Europe as an area of endemism for small mammals rather than a source for northwards postglacial colonization. – *Proc. R. Soc. B* 265: 1219–1226.
- Brunhoff, C. et al. 2003. Holarctic phylogeography of the root vole (*Microtus oeconomus*): implications for late Quaternary biogeography of high latitudes. – *Mol. Ecol.* 12: 957–968.
- Churchfield, S. and Searle, J. B. 2008. Pygmy shrew *Sorex minutus*. – In: Harris, S. and Yalden, D. W. (eds), Mammals of the British Isles: handbook. Mammal Society, pp. 267–271.
- Collins, W. D. et al. 2006. The Community Climate System Model: CCSM3. – *J. Climate* 19: 2122–2143.
- Deffontaine, V. et al. 2005. Beyond the Mediterranean peninsulas: evidence of central European glacial refugia for a temperate forest mammal species, the bank vole (*Clethrionomys glareolus*). – *Mol. Ecol.* 14: 1727–1739.
- Elith, J. and Graham, C. H. 2009. Do they? How do they? WHY do they differ? On finding reasons for differing performances of species distribution models. – *Ecography* 32: 66–77.
- Elith, J. et al. 2006. Novel methods improve prediction of species' distributions from occurrence data. – *Ecography* 29: 129–151.
- Excoffier, L. et al. 2005. Arlequin ver. 3.0: an integrated software package for population genetics data analysis. – *Evol. Bioinform. Online* 1: 47–50.
- Fløjgaard, C. et al. 2009. Ice age distributions of European small mammals: insights from species distribution modelling. – *J. Biogeogr.* 36: 1152–1163.
- Fu, Y. X. 1997. Statistical tests of neutrality of mutations against population growth, hitchhiking and background selection. – *Genetics* 147: 915–925.
- Guindon, S. and Gascuel, O. 2003. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. – *Syst. Biol.* 52: 696–704.
- Guisan, A. and Zimmermann, N. E. 2000. Predictive habitat distribution models in ecology. – *Ecol. Model.* 135: 147–186.
- Hall, T. A. 1999. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT/XP. – *Nucleic Acids Symp. Ser.* 41: 95–98.
- Hasumi, H. and Emori, S. 2004. K-1 coupled GCM (MIROC) description. – Center for Climate System Research, Univ. of Tokyo.
- Heckel, G. et al. 2005. Genetic structure and colonization processes in European populations of the common vole, *Microtus arvalis*. – *Evolution* 59: 2231–2242.
- Hewitt, G. M. 2000. The genetic legacy of the Quaternary ice ages. – *Nature* 405: 907–913.
- Hijmans, R. J. et al. 2005. Very high resolution interpolated climate surfaces for global land areas. – *Int. J. Climatol.* 25: 1965–1978.
- Ho, S. Y. W. et al. 2005. Time dependency on molecular rate estimates and systematic overestimation of recent divergence times. – *Mol. Biol. Evol.* 22: 1561–1568.
- Huelsenbeck, J. P. and Ronquist, F. 2001. MrBayes: Bayesian inference of phylogeny. – *Bioinformatics* 17: 754–755.
- Hutterer, R. et al. 2008. *Sorex minutus*. – IUCN 2009, IUCN Red List of Threatened Species, ver. 2009.2 <www.iucnredlist.org>, accessed 12 May 2009.
- Jaarola, M. and Searle, J. B. 2002. Phylogeography of field voles (*Microtus agrestis*) in Eurasia inferred from mitochondrial DNA sequences. – *Mol. Ecol.* 11: 2613–2621.
- Kotlík, P. et al. 2006. A northern glacial refugium for bank voles (*Clethrionomys glareolus*). – *Proc. Nat. Acad. Sci. USA* 103: 14860–14864.
- Librado, P. and Rozas, J. 2009. DnaSP ver. 5: a software for comprehensive analysis of DNA polymorphism data. – *Bioinformatics* 25: 1451–1452.
- Magri, D. et al. 2006. A new scenario for the Quaternary history of European beech populations: palaeobotanical evidence and genetic consequences. – *New Phytol.* 171: 199–221.
- Mascheretti, S. et al. 2003. How did pygmy shrews colonize Ireland? Clues from a phylogenetic analysis of mitochondrial cytochrome b sequences. – *Proc. R. Soc. B* 270: 1593–1599.
- McDevitt, A. D. et al. 2009. Genetic variation in Irish pygmy shrews *Sorex minutus* (Soricomorpha: Soricidae): implications for colonization history. – *Biol. J. Linn. Soc.* 97: 918–927.
- McDevitt, A. D. et al. 2010. Postglacial re-colonization of continental Europe by the pygmy shrew (*Sorex minutus*) inferred from mitochondrial and Y chromosomal DNA sequences. – In: Habel, J. C. and Assman, T. (eds), Relict species – phylogeography and conservation biology. Springer, in press.
- Mitchell-Jones, A. J. et al. 1999. The atlas of European mammals. – Poyser Natural History.
- Nogués-Bravo, D. 2009. Predicting the past distribution of species climatic niches. – *Global Ecol. Biogeogr.* 18: 521–531.
- Nylander, J. A. A. 2004. MrModeltest ver. 2. Program distributed by the author. – Evolutionary Biology Centre, Uppsala Univ.
- Ohdachi, S. D. et al. 2006. Molecular phylogenetics of soricid shrews (Mammalia) based on mitochondrial cytochrome b gene sequences: with special reference to the Soricinae. – *J. Zool.* 270: 177–191.
- Pearson, R. G. et al. 2007. Predicting species distributions from small numbers of occurrence records: a test case using cryptic geckos in Madagascar. – *J. Biogeogr.* 34: 102–117.
- Peterson, A. T. and Nakazawa, Y. 2008. Environmental data sets matter in ecological niche modeling: an example with *Solenopsis invicta* and *Solenopsis richterii*. – *Global Ecol. Biogeogr.* 17: 135–144.
- Peterson, A. T. et al. 2007. Transferability and model evaluation in ecological niche modeling: a comparison of GARP and Maxent. – *Ecography* 30: 550–560.
- Peterson, A. T. et al. 2008. Rethinking receiver operating characteristic analysis applications in ecological niche modeling. – *Ecol. Model.* 213: 63–72.
- Phillips, S. J. 2008. Transferability, sample selection bias and background data in presence-only modelling: a response to Peterson et al. 2007. – *Ecography* 31: 272–278.
- Phillips, S. J. and Dudík, M. 2008. Modeling of species distributions with Maxent: new extensions and a comprehensive evaluation. – *Ecography* 31: 161–175.
- Phillips, S. J. et al. 2004. A maximum entropy approach to species distribution modeling. – In: Greiner, R. and Schuurmans, D. (eds), Proc. 21st Int. Conf. on Machine Learning. ACM Press, pp. 655–662.

- Phillips, S. J. et al. 2006. Maximum entropy modeling of species geographic distributions. – *Ecol. Model.* 190: 231–259.
- Provan, J. and Bennett, K. D. 2008. Phylogeographic insights into cryptic glacial refugia. – *Trends Ecol. Evol.* 23: 564–571.
- Ramos-Onsins, S. E. and Rozas, J. 2002. Statistical properties of new neutrality tests against population growth. – *Mol. Biol. Evol.* 19: 2092–2100.
- Ratkiewicz, M. et al. 2002. The evolutionary history of the two karyotypic groups of the common shrew, *Sorex araneus*, in Poland. – *Heredity* 88: 235–242.
- Rogers, A. R. and Harpending, H. 1992. Population growth makes waves in the distribution of pairwise genetic differences. – *Mol. Biol. Evol.* 9: 552–569.
- Schneider, S. and Excoffier, L. 1999. Estimation of demographic parameters from the distribution of pairwise differences when the mutation rates vary among sites: application to human mitochondrial DNA. – *Genetics* 152: 1079–1089.
- Sommer, R. S. and Nadachowski, A. 2006. Glacial refugia of mammals in Europe: evidence from fossil records. – *Mammal Rev.* 36: 251–265.
- Stewart, J. R. and Lister, A. M. 2001. Cryptic northern refugia and the origins of the modern biota. – *Trends Ecol. Evol.* 16: 608–613.
- Stockwell, D. R. B. 1999. Genetic algorithms II. – In: Fielding, A. H. (ed.), *Machine learning methods for ecological applications*. Kluwer, pp. 123–144.
- Stockwell, D. R. B. and Noble, I. R. 1992. Introduction of sets of rules from animal distribution data: a robust and informative method of data analysis. – *Math. Comput. Simulat.* 33: 385–390.
- Stockwell, D. R. B. and Peterson, A. T. 2002. Effects of sample size on accuracy of species distribution models. – *Ecol. Model.* 148: 1–13.
- Svendsen, J. I. et al. 2004. Late Quaternary ice sheet history of northern Eurasia. – *Quat. Sci. Rev.* 23: 1229–1271.
- Svenning, J.-C. et al. 2008. Glacial refugia of temperate trees in Europe: insights from species distribution modeling. – *J. Ecol.* 96: 1117–1127.
- Swofford, D. L. 2000. PAUP*. Phylogenetic analysis using parsimony (*and other methods), ver. 4. – Sinauer.
- Taberlet, P. et al. 1998. Comparative phylogeography and postglacial colonization routes in Europe. – *Mol. Ecol.* 7: 453–464.
- Tajima, F. 1989. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. – *Genetics* 123: 585–595.
- Waltari, E. et al. 2007. Locating Pleistocene refugia: comparing phylogeographic and ecological niche model predictions. – *PLoS One* 7: e563.
- Willis, K. J. and van Andel, T. H. 2004. Trees or no trees? The environments of central and eastern Europe during the Last Glaciation. – *Quat. Sci. Rev.* 23: 2369–2387.
- Willis, K. J. et al. 2000. The full-glacial forests of central and south-eastern Europe. – *Quat. Res.* 53: 203–213.
- Wisz, M. S. et al. 2008. Effects of sample size on the performance of species distribution models. – *Divers. Distrib.* 14: 763–773.
- Yannic, G. et al. 2008. A new perspective on the evolutionary history of western European *Sorex araneus* group revealed by paternal and maternal molecular markers. – *Mol. Phylogenet. Evol.* 47: 237–250.