Multiple refugia and barriers explain the phylogeography of the Valais shrew, *Sorex antinorii* (Mammalia: Soricomorpha)

GLENN YANNIC1,2*, LOÏC PELLISSIER1, SYLVAIN DUBEY1,3, RODRIGO VEGA4, PATRICK BASSET5, STEFANO MAZZOTTI6, ELENA PECCHIOLI7, CRISTIANO VERNESI7, HEIDI C. HAUFFE7, JEREMY B. SEARLE4 and JACQUES HAUSSE1

1Department of Ecology and Evolution, Biophore Building, University of Lausanne, 1015 Lausanne, Switzerland
2Département de Biologie & Centre d’Études Nordiques, University Laval, 1045 Avenue de la Médecine, Québec (QC), G1V 0A6, Canada
3Shine Laboratory, School of Biological Sciences, A08, University of Sydney, Sydney, NSW 2006, Australia
4Department of Ecology and Evolutionary Biology, Corson Hall, Cornell University, Ithaca, NY 14853-2701, USA
5Hospital Preventive Medicine, University Hospital of Lausanne (CHUV), Lausanne, Switzerland
6Museo di Storia Naturale, Via Filippo de Pisis, 24, 44100 Ferrara, Italy
7Fondazione E. Mach, Research and Innovation Centre, Department of Biodiversity and Molecular Ecology, Via E. Mach 1, S. Michele all’Adige, Trento, Italy

Received 6 July 2011; revised 23 October 2011; accepted for publication 23 October 2011

The aim of the present study was to investigate the genetic structure of the Valais shrew (*Sorex antinorii*) by a combined phylogeographical and landscape genetic approach, and thereby to infer the locations of glacial refugia and establish the influence of geographical barriers. We sequenced part of the mitochondrial cytochrome *b* (*cyt b*) gene of 179 individuals of *S. antinorii* sampled across the entire species’ range. Six specimens attributed to *S. arunchi* were included in the analysis. The phylogeographical pattern was assessed by Bayesian molecular phylogenetic reconstruction, population genetic analyses, and a species distribution modelling (SDM)-based hindcasting approach. We also used landscape genetics (including isolation-by-resistance) to infer the determinants of current intra-specific genetic structure. The phylogeographical analysis revealed shallow divergence among haplotypes and no clear substructure within *S. antinorii*. The starlike structure of the median-joining network is consistent with population expansion from a single refugium, probably located in the Apennines. Long branches observed on the same network also suggest that another refugium may have existed in the north-eastern part of Italy. This result is consistent with SDM, which also suggests several habitable areas for *S. antinorii* in the Italian peninsula during the LGM. Therefore *S. antinorii* appears to have occupied disconnected glacial refugia in the Italian peninsula, supporting previous data for other species showing multiple refugia within southern refugial areas. By coupling genetic analyses and SDM, we were able to infer how past climatic suitability contributed to genetic divergence of populations. The genetic differentiation shown in the present study does not support the specific status of *S. arunchi*. © 2012 The Linnean Society of London, *Biological Journal of the Linnean Society*, 2012, 105, 864–880.


*Corresponding author. E-mail: glenn.yannic@gmail.com
INTRODUCTION

Understanding the factors that both determine the distribution of species and contribute to the formation and the maintenance of population genetic structure is a central tenet of biogeography. Moreover, such an understanding enables the prediction of the consequences of global change, such as future range contraction and loss of genetic variation. In Europe, the current patterns of species richness and genetic structure can partially be explained by constraints imposed during the Pleistocene ice ages. In particular, the three southern European peninsulas (Iberian, Italian, and Balkan) have traditionally been recognized as glacial refugia during these ice ages, and are currently considered as species-rich areas, as well as hotspots of intra-specific diversity (Bilton et al., 1998; Hewitt, 2000; Petit et al., 2003; Ruedi et al., 2008). Although the southern European peninsulas are often assumed to have been single areas from which species recolonized higher latitudes after the Last Glacial Maximum (LGM; Hewitt, 2000; Petit et al., 2003), it has recently been suggested that populations within species in a single southern peninsula may have been distributed among multiple disconnected refugia (Gómez & Lunt, 2007). Evidence for multiple glacial refugia within single southern peninsulas is now substantial (Schmitt et al., 2006; Canestrelli, Cimmaruta & Nascetti, 2007; Gómez & Lunt, 2007; Kryštufek et al., 2007; Canestrelli & Nascetti, 2008; Ruedi et al., 2008; Centeno-Cuadros, Delibes & Godoy, 2009; Grill et al., 2009). However, further studies are required to determine whether this is a common pattern or only applicable to a few species, given its importance for our interpretation of European phylogeography and our understanding of biological and genetic diversity. For example, although patterns in current genetic variation in a certain species may reflect past population structure at the LGM (i.e. the ‘refugia within refugia’ considered above; Gómez & Lunt, 2007), the patterns could also be explained by current genetic discontinuity as a result of strong geographical barriers. Therefore, the causes of genetic structure should be investigated using multiple approaches, including both species distribution modelling (SDM; Guisan & Zimmermann, 2000; Waltari et al., 2007) and the use of current landscape features to infer which factor(s) are most responsible for shaping intra-specific genetic subdivision.

SDM uses species occurrences and environmental (usually climatic) data to estimate the range of suitable environmental conditions for the species (Guisan & Zimmermann, 2000; Pellissier et al., 2010) (i.e. its environmental niche). The defined environmental niche can then be used to identify areas where the past climatic environment was suitable for the species (Nogues-Bravo, 2009), in this case at the LGM. The major advantage of such integrative approaches is that hindcasted models can be used to derive hypotheses concerning species distribution, which can subsequently be compared with the observed genetic structure (Knowles, Carstens & Keat, 2007; Richards, Carstens & Knowles, 2007).

Although the aforementioned phylogeography investigates the historical processes generating patterns of genetic variation, current landscape features, especially across increasingly fragmented landscapes, can also deeply influence the genetic diversity partitioning and gene flow between populations (Manel et al., 2003; Storfer et al., 2007). Genotyping can be combined with spatially explicit data of landscape structure (LS) and a variety of statistical methods can be used to evaluate the role that current landscape variables play in shaping current population structure and genetic diversity (Storfer et al., 2007).

The present study aimed to better understand the factors determining the current pattern of genetic variation of the Valais shrew, Sorex antinorii, over its entire distribution by combining these two approaches in an investigation of genetic structure using both SDM (hindcasted in the LGM) and LS. Despite their potential, very few studies have interrogated putative ‘refugia within refugia’ using these complementary approaches (Waltari et al., 2007). In addition, we used a framework to determine the influence of LS on current gene flow among S. antinorii populations. Sorex antinorii is a small insectivorous species belonging to the Sorex araneus group (Hoffmann, 1971; Meylan & Hausser, 1973). Its current known distribution is restricted to Italy, southern Switzerland up to the central Alps and south-eastern France (Brüner et al., 2002). It was formerly considered a chromosome race of S. araneus, although Brüner et al. (2002) argued that morphological, karyotypic, and genetic differences warrant recognizing S. antinorii as a separate species. Sorex antinorii probably diverged from the other taxa of the S. araneus group during the late Pleistocene glaciations (Taberlet, Fumagalli & Hausser, 1994; Brüner et al., 2002; Yannic, Basset & Hausser, 2008a) in refugia situated in Italy, when the Alps were covered with an immense ice sheet (Hewitt, 1996). Previous molecular studies focused in Switzerland suggested that two already differentiated genetic lineages colonized the Swiss Alps from Italy after the last glaciations, and came into secondary contact in the Rhône Valley (Lugon-Moulin & Hausser, 2002; Basset, Yannic & Hausser, 2006; Yannic, Basset & Hausser, 2008b). Mitochondrial DNA (mtDNA) has become a powerful tool for identifying evolutionary lineages or species in animals (Hebert et al., 2003; Tautz et al., 2003; Knowles & Carstens, 2007) as a result of its low or
absent recombination, uniparental inheritance, conserved structure, and relatively high evolutionary rate (Avise, 2000). The cytochrome b (cyt b) gene is one of the most frequently employed mtDNA genes for investigating phylogeographical patterns and histories at the intraspecies level. In the present study, we used mtDNA cyt b sequence data to examine the phylogeography of *S. antinorii* over its range.

**MATERIAL AND METHODS**

**PHYLOGEOGRAPHICAL ANALYSIS**

**Sampling**

The geographical locations of sampling sites are shown in Figure 1 and deposited in the Dryad repository (Yannic *et al*., 2012). For the present study, we analyzed 179 individuals from 39 localities spanning the entire known species range (i.e. Italy, France, and Switzerland) (Fig. 1). This set of samples included material collected during fieldwork and from museum collections (see Acknowledgements). Additionally, based on the results of allozymic, morphologic and morphometric studies, the existence of a relic of the subgenus *Sorex* in north-eastern Italy has been suggested (Lapini & Testone, 1998; Lapini, Filippucci & Filacorda, 2001). This taxon, named *Sorex arunchi*, was assumed to have recently diverged from *S. antinorii* (end of Pleistocene-lower Holocene) with a current occurrence in the wet lowland woods of north-eastern Italy (Terra Typica: ‘Bosco Baredi-Selva di Arvonchi’ and ‘Bosco Coda di Manin’, community of Muzzana del Turgnano, province of Udine, north-eastern Italy) (Lapini *et al*., 2001). However, no study has subsequently confirmed the existence of the taxon either genetically or karyotypically, nor established its relationship with other species of the *S. araneus* group. Therefore, six samples attributed to *S. arunchi* (Lapini *et al*., 2001) were also analyzed. Six further samples were included in the study: *S. araneus* (*N* = 2) as a sibling species of *S. antinorii* (Brüuner *et al*., 2002), *Sorex samniticus* (*N* = 2) as a sister species of the *S. araneus* group and endemic to the Italian peninsula (Fumagalli *et al*., 1996), and *Sorex minutus* (*N* = 2), which is more distantly related (Fumagalli *et al*., 1999; Yannic *et al*., 2008a, 2010) and used as the outgroup.

**DNA extraction and amplification of cyt b**

Genomic DNA was extracted using the QIAgen DNeasy Blood and Tissue kit. Double-stranded DNA amplifications of cyt b were performed with L14841/H15915 (Kocher *et al*., 1989; Irwin, Kocher & Wilson, 1991) or with a combination of primers L14841/cyt b-4 and cyt b-1/H15915 (Cyt b-1: 5’-TTC GCA GTA ATA GCC ACT GC-3’; Cyt b-4: 5’-AAC TGT TGC GCC TCA AAA TGA TAT TTG TCC TCA-3’; modified from Dubey *et al*., 2006b). Polymerase chain reactions (PCRs) were performed in a PE9700 thermal cycler (Applied Biosystems) with the cycling conditions: initial denaturation at 95 °C for 5 min, followed by 35 cycles of 94 °C for 30 s, annealing at 60 °C for 1 min and extension at 72 °C for 1 min 30 s, and a final extension of 72 °C for 10 min. The PCR products were checked on a 1.5% agarose gel and then purified using the QIAquick PCR Purification Kit in accordance with the manufacturer’s instructions. DNA sequencing was performed in a total volume of 10 µL containing approximately 100 ng of purified PCR product, 1 µL of 10 µM primers, and 4 µL of ABIPRISM Terminator 3.1 (Applied Biosystems). The sequence reaction consisted of 35 cycles of 96 °C for 15 s, 50 °C for 15 s, and 60 °C for 2 min. Purification of PCR products was conducted with a commercial kit (Qiagen) and purified PCR products were sequenced in both directions (Centre of Integrative Genomic, University of Lausanne and Cornell University Core Laboratories Center).

**Phylogenetic methods**

Nucleotide sequences of cyt b were edited in SEQUENCER, version 3.0 (Gene Codes Corp.), aligned with CLUSTALX, version 2.0 (Thompson *et al*., 1997) using default parameters, and then checked by eye and collapsed into haplotypes using DNASP, version 5.10.00 (Librado & Rozas, 2009). For Bayesian phylogenies, the best model of DNA substitution was determined using JMODELTEST, version 0.1.1 (Posada, 2008) under the Akaike information criterion. Markov Chain Monte Carlo (MCMC) technique was performed in MrBayes, 3.1.2, using a full partition strategy (i.e. each codon position for each coding gene was entered in a separate partition) (Huelsenbeck & Ronquist, 2001; Ronquist & Huelsenbeck, 2003) to characterize the probability distribution of phylogenetic trees given the data. Two independent runs were performed, each consisting of four parallel MCMC
Genetic and statistical analysis

Standard sequence polymorphism indices [number of haplotypes (Nh), polymorphic sites and parsimony informative sites] and molecular diversity indices [i.e. gene diversity (h) and nucleotide diversity (π), which are equivalent to heterozygosity at the haplotype and nucleotide level, respectively] were estimated using ARLEQUIN, version 3.5.1.2 (Excoffier & Lischer, 2010). Populations in refugial regions often show high allelic diversity as a result of refugia persistence and the accumulation of variation (Hewitt, 1996, 2001). Diversity indices were therefore estimated for the whole dataset and for the 21 sampling localities for which data on at least four specimens were available (Fig. 1, Table 1). The prediction is that a refugial population spreading from its leading edge will experience a series of bottlenecks that will reduce diversity. Thus, mtDNA diversity should decrease with distance from a refugium. This prediction has been modelled by computer simulations (Hewitt, 1996) and observed empirically (Shafer, Côté & Colman, 2011).

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 (Harpending et al., 1998). 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 using a parametric bootstrap approach as a sum of square deviations (SSD) between the observed and the expected mismatch (Schneider & Excoffier, 1999) using ARLEQUIN, version 3.5.1.2 (10,000 replicates). Fu’s (1997) Fs and Tajima’s (1989) D-tests for population expansion were performed in ARLEQUIN using coalescent simulations to test for statistical significance (10,000 replicates).

Species distribution modelling

We used records of S. antinorii throughout its range either from our own fieldwork, from databases (Centre Suisse de Cartographie de la Faune, Neuchâtel; Maiorano, Falucci & Boitani, 2008) or from museum specimens obtained for our study (see Acknowledgements; see also the Supporting information, Fig. S1). Because the occurrences were highly aggregated in some areas as a result of trapping intensity, we randomly selected a subset of occurrences with a minimal distance of 10 km. Because most modelling techniques require information about both presence and absence to determine the suitable conditions for a given species, we selected 10,000 pseudo-absences randomly over the study area covering the whole Italian peninsula, as well as the Alps; these correspond to the raw boundaries of the range occupied by the species. The modelling techniques then discriminate between the conditions suitable for presence and the background environment (Wisz & Guisan, 2009). The resulting presences and pseudo-absences were used in the subsequent SDM.

We ran species distribution models at a resolution of 2.5 arc-minutes (5 km at the equator) using eight climatic variables taken from Worldclim (Hijmans et al., 2005), expected to have a biological meaning for the distribution of S. antinorii: annual mean temperature (bio1), temperature seasonality (bio4), maximum temperature of the warmest month (bio5), minimum temperature of the coldest month (bio6), annual precipitation (bio12), precipitation of the wettest month (bio13), precipitation of the driest month (bio14), and precipitation seasonality (bio15).

We modelled the distribution of the species using the BIOMOD package (Thuiller et al., 2009), implemented for R software (R Development Core Team, 2008). Ensemble forecasting approaches have...
Table 1. Mitochondrial cytochrome b diversity statistics for 21 sampling localities of Valais shrew

<table>
<thead>
<tr>
<th>Pop ID</th>
<th>Locality</th>
<th>Country</th>
<th>Latitude</th>
<th>Longitude</th>
<th>Altitude (m)</th>
<th>N</th>
<th>Nh</th>
<th>h ± SD</th>
<th>π ± SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Méribel-Les Allues, Savoie</td>
<td>France</td>
<td>45°25’54.76&quot;N</td>
<td>06°33’22.49&quot;E</td>
<td>1109</td>
<td>19</td>
<td>12</td>
<td>0.92 ± 0.04</td>
<td>0.0038 ± 0.0023</td>
</tr>
<tr>
<td>2</td>
<td>Pralognan-La-Vanoise, Savoie</td>
<td>France</td>
<td>45°22’52.02&quot;N</td>
<td>06°43’17.86&quot;E</td>
<td>1429</td>
<td>5</td>
<td>4</td>
<td>0.90 ± 0.16</td>
<td>0.0024 ± 0.0018</td>
</tr>
<tr>
<td>3</td>
<td>Tournoux – St Paul, Alpes de Haute Provence</td>
<td>France</td>
<td>44°29’17.35&quot;N</td>
<td>06°44’22.05&quot;E</td>
<td>1480</td>
<td>15</td>
<td>6</td>
<td>0.79 ± 0.08</td>
<td>0.0030 ± 0.0023</td>
</tr>
<tr>
<td>4</td>
<td>Fromagerie, Aosta Valley</td>
<td>Italy</td>
<td>45°51’14.40&quot;N</td>
<td>07°08’60.00&quot;E</td>
<td>1631</td>
<td>6</td>
<td>4</td>
<td>0.87 ± 0.13</td>
<td>0.0024 ± 0.0018</td>
</tr>
<tr>
<td>5</td>
<td>Saint-Rhemy-En-Bosses, Aosta Valley</td>
<td>Italy</td>
<td>45°50’09.64&quot;N</td>
<td>07°11’01.22&quot;E</td>
<td>1631</td>
<td>6</td>
<td>5</td>
<td>0.93 ± 0.12</td>
<td>0.0020 ± 0.0015</td>
</tr>
<tr>
<td>6</td>
<td>St-Bernard, Valais</td>
<td>Switzerland</td>
<td>45°54’03.66&quot;N</td>
<td>07°11’42.42&quot;E</td>
<td>1920</td>
<td>8</td>
<td>5</td>
<td>0.86 ± 0.11</td>
<td>0.0026 ± 0.0018</td>
</tr>
<tr>
<td>7</td>
<td>Condove, Torino, Piemonte</td>
<td>Italy</td>
<td>45°04’39.55&quot;N</td>
<td>07°13’39.15&quot;E</td>
<td>746</td>
<td>6</td>
<td>5</td>
<td>0.93 ± 0.12</td>
<td>0.0020 ± 0.0015</td>
</tr>
<tr>
<td>8</td>
<td>Gran Paradiso National Park, Aosta Valley</td>
<td>Italy</td>
<td>45°15’09.50&quot;N</td>
<td>07°27’15.62&quot;E</td>
<td>746</td>
<td>4</td>
<td>2</td>
<td>0.67 ± 0.31</td>
<td>0.0027 ± 0.0024</td>
</tr>
<tr>
<td>9</td>
<td>Herens, Valais</td>
<td>Switzerland</td>
<td>46°02’52.99&quot;N</td>
<td>07°29’18.89&quot;E</td>
<td>1808</td>
<td>12</td>
<td>7</td>
<td>0.83 ± 0.10</td>
<td>0.0020 ± 0.0013</td>
</tr>
<tr>
<td>10</td>
<td>Lötschental, Valais</td>
<td>Switzerland</td>
<td>46°23’36.90&quot;N</td>
<td>07°45’32.54&quot;E</td>
<td>1382</td>
<td>4</td>
<td>4</td>
<td>1.0 ± 0.18</td>
<td>0.0030 ± 0.0023</td>
</tr>
<tr>
<td>11</td>
<td>Ancien Hospice, Valais</td>
<td>Switzerland</td>
<td>46°14’06.04&quot;N</td>
<td>08°00’48.88&quot;E</td>
<td>2014</td>
<td>4</td>
<td>3</td>
<td>0.83 ± 0.22</td>
<td>0.0025 ± 0.0020</td>
</tr>
<tr>
<td>12</td>
<td>Brig, Valais</td>
<td>Switzerland</td>
<td>46°16’30.70&quot;N</td>
<td>08°02’20.52&quot;E</td>
<td>1775</td>
<td>5</td>
<td>4</td>
<td>0.90 ± 0.16</td>
<td>0.0019 ± 0.0013</td>
</tr>
<tr>
<td>13</td>
<td>Tunnel Simplon, Valais</td>
<td>Switzerland</td>
<td>46°15’07.20&quot;N</td>
<td>08°02’56.40&quot;E</td>
<td>2014</td>
<td>4</td>
<td>4</td>
<td>1.0 ± 0.18</td>
<td>0.0035 ± 0.0027</td>
</tr>
<tr>
<td>14</td>
<td>Simplon, Valais</td>
<td>Switzerland</td>
<td>46°11’44.27&quot;N</td>
<td>08°03’18.98&quot;E</td>
<td>1495</td>
<td>4</td>
<td>2</td>
<td>0.67 ± 0.31</td>
<td>0.0026 ± 0.0024</td>
</tr>
<tr>
<td>15</td>
<td>Haslital, Bern</td>
<td>Switzerland</td>
<td>46°33’41.52&quot;N</td>
<td>08°20’16.24&quot;E</td>
<td>2169</td>
<td>5</td>
<td>2</td>
<td>0.60 ± 0.17</td>
<td>0.0006 ± 0.0006</td>
</tr>
<tr>
<td>16</td>
<td>Medels im Rheinwald, Graubünden</td>
<td>Switzerland</td>
<td>46°32’51.36&quot;N</td>
<td>09°17’41.24&quot;E</td>
<td>563</td>
<td>4</td>
<td>1</td>
<td>0.0 ± 0.0</td>
<td>0.0 ± 0.0</td>
</tr>
<tr>
<td>17</td>
<td>Prasöta, Sondrio, Lombardy</td>
<td>Italy</td>
<td>46°15’36.45&quot;N</td>
<td>10°15’41.55&quot;E</td>
<td>675</td>
<td>14</td>
<td>8</td>
<td>0.90 ± 0.06</td>
<td>0.0035 ± 0.0021</td>
</tr>
<tr>
<td>18</td>
<td>Lamar, Trentino-Alto Adige</td>
<td>Italy</td>
<td>46°07’44.24&quot;N</td>
<td>11°03’40.38&quot;E</td>
<td>720</td>
<td>4</td>
<td>3</td>
<td>0.83 ± 0.22</td>
<td>0.0045 ± 0.0033</td>
</tr>
<tr>
<td>19</td>
<td>Sant’Agostino, Ferrara, Emilia-Romagna</td>
<td>Italy</td>
<td>44°47’35.86&quot;N</td>
<td>11°23’21.20&quot;E</td>
<td>15</td>
<td>13</td>
<td>11</td>
<td>0.97 ± 0.04</td>
<td>0.0029 ± 0.0018</td>
</tr>
<tr>
<td>20</td>
<td>Archetón, Treviso, Veneto</td>
<td>Italy</td>
<td>46°03’51.83&quot;N</td>
<td>12°24’02.37&quot;E</td>
<td>1050</td>
<td>13</td>
<td>11</td>
<td>0.97 ± 0.04</td>
<td>0.0060 ± 0.0034</td>
</tr>
<tr>
<td>21</td>
<td>Gran Sasso National Park, Teramo, Abruzzo</td>
<td>Italy</td>
<td>42°30’32.47&quot;E</td>
<td>13°30’45.99&quot;E</td>
<td>1427</td>
<td>9</td>
<td>6</td>
<td>0.89 ± 0.09</td>
<td>0.0040 ± 0.002</td>
</tr>
</tbody>
</table>

Included are the number of individuals in each site (N), observed haplotypes (Nh), haplotype diversity (h), and nucleotide diversity (π) with their standard deviation (SD).
been shown to significantly improve the accuracy of species distribution models (Marmion et al., 2009). Therefore, we used and combined the results of seven different statistical techniques to model the distribution of the species: (1) generalized linear model (GLM); (2) generalized additive model (GAM); (3) classification tree analysis (CTA); (4) artificial neural networks (ANN); (5) multivariate adaptive regression splines (MARS); (6) generalized boosting model (GBM); and (7) Random Forest (RF).

To evaluate the predictive performance of the species distribution model, we used a random subset of 70% of the data to calibrate every model, and used the remaining 30% for the evaluation. Models were evaluated using a relative operating characteristic (ROC) curve and the area under the curve (AUC) (Fielding & Bell, 1997). We repeated the split 50 times and recalculated the average AUC of the repeated split-samples, which gave a more robust estimate of the predictive performance of each model.

Finally, each model was projected into the past using two general circulation model (GCM) simulations for the last glacial maximum (LGM: 21 000 ± 2000 years): the Worldclim data of the Community Climate System Model (CCSM; Collins et al., 2004) and the Model for Interdisciplinary Research on Climate (MIROC, version 3.2; Hasumi & Emori, 2004) downscaled to a resolution of 2.5 (4 km) arc-minutes resolution. To reflect the central tendency of these distributions, accounting for variations among modelling techniques, we applied a weighted average of the seven modelling techniques based on the predictive power (AUC; Araújo & New, 2007). Predictions of species distributions were obtained by classifying the probabilities into binary presence and absence data according to a ROC-optimized threshold, which is considered among the best-performing threshold-based approaches (Liu et al., 2005).

**LANDSCAPE DATA AND LANDSCAPE RESISTANCE MODELS**

Although mtDNA evolves too slowly to be useful for inferring most recent and ongoing micro-evolutionary processes, the variations in haplotype frequencies are still informative for identifying landscape processes shaping genetic structure through gene flow (Wang, 2010).

For this analysis, the dataset was reduced to the sampling localities for which data on at least four specimens were available, and we excluded the southernmost population Gran Sasso, Abruzzo, because its distance from the others exceeded computational limitations. Therefore, the landscape analysis included 154 out of the 179 S. antinorii individuals and 89 out of the 103 inferred mtDNA haplotypes from 20 different sampling localities (Fig. 1, Table 1). The population structure across the study area and between sampling sites was assessed by calculating Fst, using ARLEQUIN. For the genetic model, we used the Kimura two-parameter genetic distance (Kimura, 1980). Significance values for the two methods of computation of population structure were obtained after 10 000 permutations.

We used CIRCUITSCAPE, version 2.2 (McRae, 2006) to model the connectivity between populations accounting for landscape features, which can enhance or limit the dispersal of S. antinorii. The algorithm in CIRCUITSCAPE evaluates landscape resistance or conductance between the investigated populations from multiple paths (McRae, 2006). For this analysis, we first generated a raster of landscape resistance based on a ‘flat’ landscape (i.e. all pixels with the same resistance value) at a resolution of 300 m for an area containing the 20 populations analyzed. Second, we generated a digital elevation model (DEM) at a resolution of 300 m from the raster DEM. Because it is expected to be more costly for the species to climb to a higher elevation to disperse, higher altitude can be seen as a resistance to connectivity. Third, because the species is known to use moist (i.e. riverside used as corridor) habitat with a dense vegetation cover (Lugon-Moulin, 2004), we created a raster of distances to rivers, assigning pixels far from rivers as more resistant to dispersal (RIV). These two landscape rasters were rescaled to have values between 0 and 1. Finally, we extracted a raster of land cover (LAC) from the ESA-GlobCover at a resolution of 300 m. We assigned conductance values from 0 to 1 to each land cover categories using our knowledge of the ecology of the species (Lugon-Moulin et al., 1999; Lugon-Moulin & Hausser, 2002; Lugon-Moulin, 2004; Yannic et al., 2008b) and expert knowledge from the literature (Murray et al., 2009). All large water bodies were given a conductance a priori of 0 in all rasters. We also generated input rasters for CIRCUITSCAPE combining pairs of landscape features corresponding to the sum of the rasters previously calculated (McRae & Beier, 2007). We generated a pairwise connectivity matrix based on the four rasters above and their combination. To evaluate the relative importance of the landscape features in predicting levels of genetic structure across the population studied, we conducted Mantel (1967) tests examining correlations between pairwise genetic structure and models of pairwise connectivity. All Mantel tests were conducted in the R software package ECODIST (Goslee & Urban, 2007) with 10 000 matrix permutations to assess significance.

RESULTS

PHYLOGEOGRAPHICAL ANALYSIS AND MOLECULAR DATING

A total of 103 haplotypes was identified among the 185 specimens and deposited in GenBank (accession numbers: HQ901808–HQ901910). Of the 1011 bp sequenced, 115 sites were variable and 57 parsimony-informative. No insertions or deletions were observed. The average transitions/transversion ratio (5.5) and base composition (T: 29.2%; C: 28.2%; A: 28.8%; G: 13.8%) are similar to values reported in previous studies of the cyt b gene of several small mammals (Michaux et al., 2003; Deffontaine et al., 2005).

The overall observed gene diversity (\(h\)) was 0.972 ± 0.008 (mean ± SD) and the overall nucleotide diversity (\(\pi\)) was 4.9 × 10\(^{-3}\) ± 0.24 × 10\(^{-3}\). Gene diversity within sampling sites ranged from 0.0 to 1.0 (median = 0.89) and nucleotide diversity varied from 0.0 to 6.0 × 10\(^{-3}\) (median = 2.6 × 10\(^{-3}\)). Figure 2 revealed a higher nucleotide diversity in north-eastern Italy, whereas gene diversity is rather homogeneous among sampling sites. There was a significant and negative correlation between the nucleotide diversity and Euclidian distance to the most eastern population (i.e. Archeton, Treviso: \(r = -0.44, P = 0.043\)), and this effect is even stronger when the monomorphic population (i.e. Medels im Rheinwald, Graubünden, Switzerland) was removed from the analysis (\(r = -0.60, P = 0.005\)). There was no correlation between gene diversity and Euclidian distance to the most eastern population (\(r = 0.007, P = 0.98\)).

The mismatch distribution of the whole dataset showed a unimodal distribution that fitted, visually, almost perfectly over the expected values for a population expansion model (data not shown). There was an observed mean of 4.98 ± 2.43 pairwise differences among haplotypes. The goodness-of-fit test showed no significant differences between the observed and expected values under a sudden expansion model (SSD = 0.0001, \(p_{SSD} > 0.05\); \(rg = 0.0074, p_{rg} > 0.05\)). Negative and significant Tajima’s \(D\) (−2.3509, \(P < 0.001\)) and Fu’s \(F_s\) (−25.2442, \(P < 0.001\)) showed departures from neutrality also consistent with a sudden population expansion.

Bayesian phylogenetic analyses inferred with a HKY+G+I model revealed limited support for phylogenetic structure within \(S.\) antinorii because they are essentially polytomies (Fig. 3). Several statistically supported haplogroups emerged, essentially composed of samples found at the margin of the \(S.\) antinorii range (i.e. mostly located in eastern Alps but also in western and north-western Alps or in the Apennines). The six samples attributed to \(S.\) arunchi fell into to the main lineage and did not differ from those of \(S.\) antinorii (Fig. 3, black stars).

The haplotype network displays a star-like pattern with a central high-frequency haplotype (\(f = 0.15\)) (Fig. 4). The central common haplotype was found in Italian, Swiss, and French localities, although not in the north-eastern Italian localities. Three clusters emerged, which were also supported on the BI tree. Interestingly, all these three divergent haplogroups are located in central or eastern Alps. In agreement with the BI analyses, three additional haplogroups, less distant to the central haplotype, are also present. They encompass samples found in western Alps, as
well as in the Apennines. Again, the six samples attributed to *S. arunchi* did not differ from those of *S. antinorii*.

**Species distribution modelling**

SDM proved useful for predicting the distribution of *S. antinorii* (AUC: ANN = 0.776, CTA = 0.831, GAM = 0.856, GBM = 0.846, GLM = 0.832, MARS = 0.828, RF = 0.84). The overall results show that the potential distribution for *S. antinorii* (estimated using recent species records and eight selected bioclimatic variables) encompasses the known distribution of the species in Europe (Fig. 5A). However, the model also found suitable habitat for *S. antinorii* outside its actual range or where the species has not yet been recorded despite extensive sampling efforts (i.e. west of the French Rhône Valley and in the Jura Mountains).

The two GCMs predicted fragmented suitable LGM climatic conditions for *S. antinorii* in the Italian Peninsula, concordant with distinct refugia within a refugium (Fig. 5B, C). The potential niche predicted under the MIROC model (Fig. 5C) is generally more fragmented and restricted than the CCSM predicted distribution (Fig. 5B). Both GCMs, however, predicted patchy suitable LGM climatic conditions in an extended area, ranging from the Region of Piedmont to the Apennines of the Region of Abruzzo, and also to the Region of Calabria on the southern tip of the peninsula. CCSM and MIROC also predicted more restricted suitable habitats close to the edge of the ice sheet present in north-eastern Italy during the LGM. Nonetheless, suitable LGM climatic conditions were also predicted outside the Italian Peninsula by the two models: (1) east of Italy, in the Balkans, on the eastern coast of the Adriatic Sea; (2) west of Italy, in the French-Italian Alps in south-eastern France.

**Figure 3.** Bayesian phylogeny for *Sorex antinorii*, *Sorex arunchi*, *Sorex araneus*, and *Sorex samniticus*. The genealogy is based on cyt *b* gene haplotypes (1011 bp) and rooted with *Sorex minutus*. Node labels represent posterior probabilities ≥ 0.9 (highlighted by bold branches). Black stars indicate haplotypes putatively attributed to *S. arunchi*. Geographical regions refer to the range definition of Fig. 1.
extending as far north as the Vercors Massif; and (3) north of the LGM ice sheet in the German Alps and neighbouring north-eastern areas.

**LANDSCAPE RESISTANCE MODELS**

A hierarchical analysis of molecular variance showed that most of the mtDNA variability (65%) was distributed within populations. The overall genetic differentiation of populations was high and significant ($\Phi_{ST} = 0.35$, $P < 0.001$). Pairwise genetic distances between sampling localities ranged from zero to 0.65. We observed a significant pattern of isolation-by-resistance (IBR) based on the ‘flat’ landscape at this spatial scale ($R^2 = 0.20$, $P < 0.0001$) (Table 1). By comparison, IBD based on Euclidean distances between sites explained less variance ($R^2 = 0.17$, $P < 0.0001$). Landscape resistance values that incorporated altitude (DEM) and, to a lesser extent, land cover (LAC) as dispersal barriers resulted in a significant but stronger relationship between landscape resistance and genetic structure than those based on GEO distances or a ‘flat’ landscape (Table 2). Incorporating the distance to rivers (RIV), the model suggested that there was no significant relationship between genetic structure and geographical features, after correction for multiple tests. The incorporation of other landscape variables or combinations of other landscape variables did not further improve the relationship between genetic structure and geographical distance (Table 2).

**DISCUSSION**

**CLIMATIC SUITABILITY AT THE LGM**

The concordance between the two species distribution models suggests that we obtained robust results concerning the LGM distribution of *S. antinorii*. The two-hindcated models showed some discontinuities in the range of suitable climatic conditions for this
species at the LGM (LGM: 21 000 ± 2000 years ago), which could represent error during climatic reconstructions. The predictions for the LGM identified two main climatically suitable areas within the Italian peninsula concordant with the ‘refugia within refugia’ concept (Gómez & Lunt, 2007). The larger of these two areas was in the region of Piedmont and in the Apennine mountain chain, from the northern Apennines to Abruzzo. Interestingly, the second suitable habitat, covering a smaller area, also appeared to be present in northern Italy, at the border of the ice sheet in the pre-Alps of Lombardy. The models used predicted both present day and past suitable conditions outside the reported distribution of the species: to the west, from the Upper Arve Valley to the Vercors Massif and in the Jura Mountains; in the eastern and northern coasts of the Adriatic Sea; and to the north of the ice sheet in the German Alps. We have no evidence that confirms the presence of *S. antinorii* in these regions during the LGM. This situation probably reflects the existence of competing forms (the area concerned being currently occupied by the sibling species *S. araneus* and *S. coronatus*) and past dispersal barriers (extended glaciers). Therefore, at the LGM, *S. antinorii* was apparently restricted to the Italian Peninsula, which itself was subdivided into multiple suitable areas, concordant with the ‘refugia within refugia’ concept (Gómez & Lunt, 2007).

**PHYLOGEOGRAPHICAL APPROACH**

The phylogeographical analysis revealed shallow divergence among haplotypes and no clear substructure within *S. antinorii*. The comparison of the two tests of neutrality and the starlike topology of the median-joining network both indicated a sudden population expansion from a single refugium, probably located in the Apennines (see also the map of suitable available habitats during the LGM; Fig. 5). Furthermore, additional haplogroups, statistically supported, emerged on the Bayesian analyses and the median-joining network also showed long branches. These long branches notably lead to haplogroups located in the north-eastern part of the Italian Peninsula, suggesting that at least another refugium may have existed there. Populations in refugial regions often show high genetic diversity due to refugial persistence and accumulation of variation (Hewitt, 2001, 2004). In accordance with this prediction, we observed higher nucleotide diversity (π) in populations from eastern Alps. However, gene diversity (h) was not higher in these populations. The higher π values observed in north-eastern Italy might be explained by an intrinsic characteristic of this parameter that takes into account the divergence between haplotypes and therefore it can be inflated if haplotypes from...
different refugia meet in a zone of contact. Thus, large \( \pi \) values may be found in areas that have received immigrants from more than one refugium. Nowadays, there is no longer geographical segregation between the lineages, suggesting that the populations came into contact soon after a period of genetic differentiation.

It is worth noting that a variety of other small vertebrates, including amphibians (Canestrelli et al., 2007; Canestrelli & Nascetti, 2008), reptiles (Ursenbacher et al., 2006), and mammals (Grill et al., 2009; Vega et al., 2010) are also characterized by divergent genetic lineages within the Italian Peninsula. These previous studies also concur that the southern part of the Italian Peninsula is a particularly important site of genetic diversification, and repeatedly show phylogeographical discontinuities in the Calabrian Peninsula (Canestrelli, Cimmaruta & Nascetti, 2008; Vega et al., 2010). Our data based on genetics and SDM differ, however, from most previous studies (1) in the level of genetic differentiation among clades and (2) mainly in suggesting possible refugia located in the northern Italian Peninsula, consistent for a cold-tolerant species, such as *S. antinorii*. Such a pattern has also been documented in *Hyla intermedia*, an amphibian that forms three well-supported clades at the mtDNA level; one clade being restricted to the north of Italy (Dubey, Ursenbacher & Fumagalli, 2006a; Canestrelli et al., 2007; Stoecck et al., 2008). A similar pattern was also observed in the wall lizard (*Podarcis muralis*), where one clade is restricted to the Alps and the western Padana Plain, and the other two are located on the Tyrrhenian side of Italy, in the central Apennines and southern Italy, respectively (Giovannotti, Nisi-Cerioni & Caputo, 2010). Nevertheless, the ecology and geographical range of the different lineages inferred for both species indicate that it is unlikely that they had a similar diversification history to *S. antinorii*. Instead, the border of the southern European Alps is known to be a glacial refugium for several alpine plant species (Schönswetter et al., 2005), which suggests that there was also suitable habitat for small mammals in the pre-Alps region at the LGM. Unexpectedly, the genetic substructure previously discovered within *S. antinorii*, primarily on the basis of microsatellite analysis (i.e. one group containing individuals sampled in the northern part of the French Alps and western Switzerland and the second group containing the individuals sampled in Italy, eastern Switzerland and the southern French Alps) ( Lugon-Moulin & Hausser, 2002; Basset et al., 2006; Yannic et al., 2008b), is not geographically confirmed here at a broader geographical scale with cyt b. This substructure could therefore be the result of a regional genetic isolation of populations rather than a more ancient phylogeographical differentiation.

**Effect of the landscape**

Current landscape features, especially across increasingly fragmented habitats, can also deeply influence the partitioning of the genetic diversity and gene flow between populations (Keyghobadi, 2007; Storfer et al., 2007; Holderegger & Wagner, 2008; Holderegger & Di Giulio, 2010). Along its length, the Italian Peninsula is highly fragmented by large urban and suburban infrastructures, wide rivers, and mountainous

---

**Table 2.** Results of Mantel tests showing the association between pairwise genetic distance \( \phi_{ST}/(1 - \phi_{ST}) \) and models of geographical distance among Valais shrew populations

<table>
<thead>
<tr>
<th>Geographical variable</th>
<th>( r )</th>
<th>( R^2 )</th>
<th>95%CI</th>
<th>( P )</th>
</tr>
</thead>
<tbody>
<tr>
<td>GEO</td>
<td>0.408</td>
<td>0.166</td>
<td>0.360/0.456</td>
<td>&lt; 0.0001*</td>
</tr>
<tr>
<td>FLAT</td>
<td>0.451</td>
<td>0.20</td>
<td>0.394/0.527</td>
<td>&lt; 0.0001*</td>
</tr>
<tr>
<td>LAC</td>
<td>0.505</td>
<td>0.25</td>
<td>0.304/0.461</td>
<td>&lt; 0.0001*</td>
</tr>
<tr>
<td>DEM</td>
<td>-0.591</td>
<td>0.35</td>
<td>-0.655/-.509</td>
<td>&lt; 0.0001*</td>
</tr>
<tr>
<td>RIV</td>
<td>-0.190</td>
<td>0.036</td>
<td>-0.278/-.090</td>
<td>0.0129</td>
</tr>
<tr>
<td>LAC_RIV</td>
<td>0.368</td>
<td>0.13</td>
<td>0.304/0.461</td>
<td>&lt; 0.0001*</td>
</tr>
<tr>
<td>DEM_RIV</td>
<td>-0.372</td>
<td>0.14</td>
<td>-0.441/-.280</td>
<td>&lt; 0.0001*</td>
</tr>
<tr>
<td>DEM_LAC</td>
<td>0.225</td>
<td>0.050</td>
<td>0.168/0.310</td>
<td>0.0023*</td>
</tr>
</tbody>
</table>

\( R^2 \), the proportion of the variance explained by the model; 95%CI, the 95% confidence limits of the Mantel \( r \) and \( P \) two-tailed \( P \)-values (null hypothesis: \( r = 0 \)). Asterisks (*) indicate significant \( P \)-values, after adjustment for multiple tests, based on a sequential goodness of fit meta-test (SGoF; Carvajal-Rodriguez, de Una-Alvarez & Rolan-Alvarez, 2009).

DEM, elevation model; LAC, land cover; FLAT, ‘flat’ landscape; GEO, Euclidian distance; RIV, river (for details, see Material and Methods).
landscapes. Typically, these features are considered to impede dispersal and reduce gene flow (Trombulak & Frissell, 2000; Delaney, Riley & Fisher, 2010; Frantz et al., 2010; Murphy et al., 2010). Our IBR study also showed that some landscape features probably had an impact on the genetic differentiation among populations of *S. antinorii* when we controlled for distance between localities. In particular, we showed that altitude and land cover had a strong effect on population genetic differentiation. Although current occurrences of *S. antinorii* are recorded up to 2700 m a.s.l. (Yannic et al., 2008b) and previous studies showed that alpine passes of up to 2500 m a.s.l. did not represent strong barriers to gene flow for *S. antinorii* (Lugon-Moulin & Hausser, 2002; Yannic et al., 2008b), it is not so unexpected that glacier-covered mountain ridges and predominately rocky habitats strongly impact gene flow. Conversely, rivers apparently had no significant impact on gene flow. This result is consistent with previous studies (Lugon-Moulin et al., 1999), although such a finding may depend on the nature of the streams (mountain streams and moraine may impeded gene flow).

Our LS approach showed that heterogeneous landscape (e.g. altitude and land cover) may affect genetic differentiation among shrew populations. We have also previously demonstrated that *S. antinorii* appears to have occupied disconnected glacial refugia in the Italian peninsula during the LGM. Based on both approaches, it is however difficult to disentangle the main factors (i.e. current landscape features or past isolation during the LGM) explaining the observed current genetic differentiation of shrew populations. Two main reasons can be advocated. First, the cyt *b* is not the most suitable marker to infer current gene flow. Second, the populations used for the LS analyses are mainly located in the northern range of the species (i.e. where the putative cryptic refugia were located and where the altitudes are also the highest). Therefore, both effects may be mingled.

**Specific status of *S. arunchi*, Lapini & Testone, 1998**

*Sorex antinorii* belongs to the *S. araneus* group, encompassing nine morphologically, genetically, and chromosomally well-described species (Fumagalli et al., 1996; Searle & Wójcik, 2000; Brünner et al., 2002). *Sorex antinorii* and related species also show impressive diversification involving chromosomal rearrangements. Such variability reaches its maximum in *S. araneus*, a Palearctic species differentiated in >70 different karyotypic races (Searle & Wójcik, 1998; Wójcik et al., 2003). *Sorex arunchi* has been described on the basis of morphology and morphometrics (Lapini & Testone, 1998; Lapini et al., 2001). Describing new species from morphologically homogeneous but species-rich groups such as *S. araneus* is notoriously difficult. Analyses of standard DNA markers are often useful for resolving such taxonomic problems. Therefore, the present study included six samples attributed to *S. arunchi*, which were kindly provided by Luca Lapini (Museo Friulano di Storia Naturale, Udine) and morphologically identified. However, despite the possibility of cryptic subclades within *S. antinorii* as shown by our SDM approach, the phylogenetic positions of these samples did not allow the distinction of *S. arunchi* from *S. antinorii*. Indeed, exactly the same haplotypes were shared between the two taxa. Introgressive hybridization leading to massive transfer of mtDNA haplotypes from a species to another is not an uncommon phenomenon in mammals (Ruedi, Smith & Patton, 1997; Alves et al., 2006; Pidancier et al., 2006; Gompert et al., 2008; Good et al., 2008) and has probably occurred among species of the *S. araneus* group (Yannic et al., 2008a, 2010). Therefore, additional samples from north-eastern Italy, where genetic differentiation has most likely occurred (Figs 3, 4), as well as alternative marker systems (autosomal and Y-chromosome genes) and karyological data, are certainly required to accurately investigate *S. arunchi* properly. For now, the lack of genetic differentiation shown in the present study does not support the specific status of *S. arunchi*.

**Conclusions**

Long periods of geographical isolation during the Pleistocene glaciations are viewed as the main causes of genetic differentiation and subsequent speciation of current fauna. Our results with SDM confirmed the possibility of multiple refugia in the Italian Peninsula and the shallow divergence within *S. antinorii* may result from both historical processes (as demonstrated by phylogeographical approaches) and contemporary processes (as suggested by our IBR approach). Contrasting genetic structure inferred from mtDNA against markers with faster evolutionary rates (i.e. nuclear microsatellites) would, however, be required to fully disentangle the intricate role of historical vicariance and contemporary fragmentation that influence the distribution and abundance of genetic diversity in the Valais shrew.

**Acknowledgements**

We are grateful to everyone who provided tissue samples of shrews over the years: Patrick Brunet-Lecomte, Riccardo Castiglia, Paolo Debernardi, André Meylan, Jean-François Noblet, Elena Patriarca,
Françoise Poitevin, and Nigel G Yoccoz. We thank Simon Capt and the Centre Suisse de Cartographie de la Faune, Neuchâtel, Switzerland, for providing occurrence data of *S. antinorii* in Switzerland, as well as all the field observers: Fabien Fivaz, Jérome Fournier, Sandrine Jutzeler, Tiziano Maddalena, Paul Marchesi, Peter Vogel, and Mirko Zanini. We also thank Agnès Horn for her contribution to the laboratory work and Lucie Büchi for her graphical expertise. This study was supported by Fondation Agassiz, University of Lausanne and Société Académique Vaudoise (Switzerland) grants to G.Y. We thank the Centre for Alpine Ecology, the Fondazione Edmund Mach, and the Autonomous Province of Trento for supporting the research conducted by H.C.H., E.P., and C.V. We especially thank the three anonymous reviewers for their comments on an earlier version of the manuscript.

REFERENCES


Rambaut A. 2009. FigTree, version 1.2.2. Available at: http://tree.bio.ed.ac.uk/software/figtree/.


Stoecck M, Dubey S, Klutsch C, Litvinchuk SN, Scheidt —


SUPPORTING INFORMATION

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

**Figure S1.** Location of samples used for the species distribution modelling analyses.

Please note: Wiley-Blackwell are not responsible for the content or functionality of any supporting materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.