Phylogeographic and morphometric studies on the Eurasian

pygmy shrew *Sorex minutus*: insights into its evolutionary history

and postglacial colonisation in Europe

Rodrigo Rafael Vega Bernal

PhD Thesis

University of York

Department of Biology

July 2010

**Abstract**

Here, I investigate the phylogeography and morphology of the Eurasian pygmy shrew *Sorex minutus*, searching for significantly differentiated lineages, colonisation routes and demographic parameters that would explain the effects of the Quaternary glaciations on the current distribution of the species. I also explore the genetic and morphological diversity and origin of pygmy shrew populations in the British Isles, particularly focusing on Ireland and the Orkney islands. Mitochondrial and nuclear DNA markers were used for the phylogeographic analyses, and a geometric morphometrics approach was implemented on mandible and skull samples. There was an evident phylogeographic structure across Eurasia consistent with occurrence of southern glacial refugia, and there were two distinct lineages in Northern-Central Europe and near the Pyrenees supporting the existence of northern glacial refugia through the characteristics of their distribution and population expansion. Haplotypes from Britain belonged to these two northern lineages, with the Pyrenean lineage forming a peripheral 'Celtic fringe'. I show that it is most likely that pygmy shrews on both Ireland and Orkney were introduced by humans from mainland British Celtic fringe rather than further afield, even though there is a haplotype found in Northern Spain identical to one in Ireland. Mandible size increased noticeably with decreasing latitude, but skulls showed no evident trend in size variation. Shape variation was significant but modest when analysing the sample divided into phylogeographical groups. However, the samples from different islands within the British Isles show that island evolution played an important role in morphological diversity, with mandible and skull shape divergence on small islands and low genetic diversity. These results notably expanded previous findings and indicate that *S. minutus* is an excellent model for understanding the effects of climate change on biological diversity, colonisation and differentiation in

refugia, and island evolution, useful for the conservation of genetic and

morphological diversity.

**List of contents**

**List of tables**

**List of figures**

**Figure 1.1.** A) Map of Eurasia showing the geographical distribution of the Eurasian pygmy shrew *Sorex minutus* (modified from Corbet and Harris 1991). B) Map of Europe showing the three Mediterranean peninsulas (namely Iberian, Italian and Balkan) and the glaciated areas at the Last Glacial Maximum (gray shaded polygons delimited by thick black lines; modified from Ehlers and Gibbard 2004).

**Figure 2.1.** Sample localities of *Sorex minutus* from Europe divided into cytochrome *b* phylogroups: ● North-Central European (three Russian samples from Siberia are not shown in map), ○ Pyrenean, ☆ South Italian, □ North-Central Italian, ▽ Balkan and ▲ Iberian; ■ *S. volnuchini* (outgroup). Relief illustrated in grey tones.

**Figure 2.2.** Bayesian inference tree of cytochrome *b* data showing the phylogenetic relationships among *Sorex minutus* haplotypes. Phylogroups (North-Central European, Pyrenean, South Italian, North-Central Italian, Balkan and Iberian) and four other groups (Britain, Ireland, Orkney and Celtic fringe) are represented with different colours. Letters 'A – F' correspond to ancestral (central) haplotypes in the MJ network (Fig. 2.3). Values on branches correspond to posterior probability support. The scale bar represents 0.3 nucleotide substitutions per site.

**Figure 2.3.** Median-Joining network of cytochrome *b* data showing the relationships among *Sorex minutus* haplotypes. A) Full network showing the main phylogroups (dashed lines). B) The North-Central European phylogroup (with British group). C) The Pyrenean (with Ireland, Orkney and Celtic fringe groups), North-Central Italian and South Italian phylogroups. D) The Balkan and Iberian phylogroups. Ancestral (central) haplotypes from where all other haplotypes derive from are marked with

ix

letters 'A – F' (see text for explanation). Numbers in parentheses represent mutation steps. Dotted line in centre of A) shows where the network was cut.

**Figure 2.4.** Bayesian inference tree of A) Y-chromosome introns and B) BRCA1 gene showing the phylogenetic relationships among *Sorex minutus* sequences. Values on branches represent posterior probability branch support.

**Figure 2.5.** Neighbor-Joining tree showing the pairwise distances among cytochrome *b* phylogroups of *Sorex minutus* and estimated divergence times (in MYA) assuming a molecular clock.

**Figure 2.6.** Bayesian inference tree of cytochrome *b* data showing the genetic relationships among phylogroups of *Sorex minutus*. Values on nodes correspond to Times to the Most Recent Common Ancestor with 95% Highest Probability Density confidence intervals (in MYA). The ingroup (*) was calibrated with the oldest known fossil record of *S. minutus* (5 MYA, SD = 0.2).

**Figure 2.7.** Mismatch distribution (frequency of pairwise nucleotide differences) of cytochrome *b* phylogroups of *Sorex minutus*.

**Figure 2.8.** Bayesian Skyline Plots of cytochrome *b* phylogroups of *Sorex minutus* showing the historical demographic trends. The solid lines are median estimates and the shaded areas represent 95% Highest Probability Densities (confidence intervals).

**Figure 3.1.** Samples of *Sorex minutus* for morphological analysis. A) Mandible dataset. B) Skull dataset. The symbols correspond to the different lineages detected

(RW1 and RW2), and the average configuration are shown for purposes of visualisation.

**Figure 3.13.** Canonical Variate Analysis of shape variables of *Sorex minutus* skulls showing group differences. The first two canonical variates (CV) were used to describe variation among groups.

**Figure 3.14.** A) Unrooted and B) rooted Neighbor-Joining (NJ) trees constructed using skull Procrustes distances among *Sorex minutus* groups. Branch support is shown as bootstrap values (10,000 replications). The grids on the right correspond to shape deformation per group from the average configuration. Inset: NJ trees based on genetic distances obtained in Chapter 2, shown for purposes of comparison.

**Figure 4.1.** Samples of *Sorex minutus* for genetic analysis. ⛝ Continental, ∎ British Celtic fringe, ◯ Ireland and △ Orkney islands.

**Figure 4.2.** Median-Joining network of cytochrome *b* data showing the relationships among *Sorex minutus* haplotypes belonging to the Pyrenean lineage. The Pyrenean lineage is divided into a 'Continental' group (light green), a 'Celtic fringe' group distributed on Britain and some islands off the coast of Britain (light blue), excluding the Orkney islands, 'Ireland' (dark green) and 'Orkney islands' groups (purple). Pointed lines indicate the central (ancestral) haplotypes from where all other haplotypes derive from (see text for explanation). Some loops in the network are not on scale and mutations are shown with numbers. For the 'Continental' group, the country of origin for the haplotypes is shown (FR = France, ES = Spain).

**List of appendices.**

**Appendix 1.**  Genetic and morphological variation in a Mediterranean glacial refugium: evidence from Italian pygmy shrews, *Sorex minutus* (Mammalia: Soricomorpha).

**Appendix 2.** Northern glacial refugia for the pygmy shrew *Sorex minutus* in Europe revealed by phylogeographic analyses and species distribution modelling.

**Appendix 3.** List of samples of *Sorex minutus* for genetic analysis.

**Appendix 4.** List of mandible and skull samples for geometric morphometric analysis.

**Appendix 5.** Analysis of Variance (ANOVA) post-hoc tests on Centroid Size (CS) for *Sorex minutus* mandibles.

**Appendix 6.** Multivariate Analysis of Variance (MANOVA) post-hoc tests on shape variables for *Sorex minutus* mandibles.

**Appendix 7.** Analysis of Variance (ANOVA) post-hoc tests on Centroid Size (CS) for *Sorex minutus* skulls.

**Appendix 8.** Multivariate Analysis of Variance (MANOVA) post-hoc tests on shape variables for *Sorex minutus* skulls.

**Appendix 9.** Analysis of Variance (ANOVA) of *Sorex minutus* mandibles.

**Dedication**

To my nieces and nephews, Ximena, Alejandro, Evelyn, Oriana and Federico, and to my grand-niece Amalia.

## Acknowledgements

these years unforgettable. The time while in the US would have been less fun without my friends Guillermo, Mathieu, Marie-Line, Aldo, Patricia, Sergio and the rest of the Spanish gang.

Back in México and Argentina, Juan Pablo, Rodrigo, Alfredo, Julieta, Andrés, Giuseppe, Yoshi, Mariana and Ariel always reminded me of the joys of life. I missed you.

My parents have given me endless support, love and courage to keep going this far. I hope they enjoy their well deserved retirement. My brother and sister, uncles and cousins have also been there for me when I needed them the most.

And finally, I would like to thank my wife Sophie for giving me more than I ever thought I could get from a relationship.

**Author declaration**

I declare that this is my own work, unless stated otherwise.

# Chapter 1

# General introduction

## 1.1 Structure of the thesis

In this thesis I explore the genetic and morphological diversity of the Eurasian pygmy shrew *Sorex minutus* (Linnaeus 1766) by means of a phylogeographic and geometric morphometrics approach.

This chapter (Chapter 1) gives a general introduction to the subject species in a biogeographical, taxonomic and evolutionary setting, and I provide the aims of this study. In Chapter 2, I perform a phylogeographic analysis using mitochondrial and nuclear DNA markers, and discuss the findings in the context of the Quaternary glaciations and the processes of Eurasian colonisation and evolutionary divergence. Chapter 3 is devoted to the analysis of morphological variation using a geometric morphometric technique on mandible and skull datasets of *S. minutus* obtained across Europe. In Chapter 4, I present a combined analysis of phylogeography and geometric morphometrics on the genetic and morphological variation of *S. minutus* belonging to a Pyrenean lineage, distributed in Western continental Europe, the Orkney islands, Ireland and Britain, to understand the colonisation patterns and divergence in an island setting. In the final chapter (Chapter 5), I discuss the main findings and future directions, and present the general conclusions of my phylogeographic and morphometric studies on *S. minutus*. The appendices 1 and 2 correspond to two published works in scientific journals that originated from this thesis, one in the Biological Journal of the Linnean Society (Vega et al. 2010a) and the other in Ecography (Vega et al. 2010b), made in collaboration with other researchers; the former explores the genetic and morphological distinction of *S.*

*minutus* in the Italian peninsula, while the later examines the northern glacial refugia hypothesis with a phylogeographic and ecological niche modelling approach.

**1.2 The European peninsula as a geographical setting**

Europe is a macropeninsula with an east-west orientation on the western-most region of the Eurasian landmass forming part of the Palaearctic biogeographic region. Europe is delimited in the east by the Ural Mountains, in the southeast by the Caspian Sea, Caucasus Mountains and the Black Sea, in the south by the Mediterranean Sea, in the west by the Atlantic Ocean and in the north by the Arctic Ocean (Fig. 1.1A, B; Temple and Terry 2007). The total area (including European Russia) is more than 10 million square kilometres, and the most important mountain ranges are the Pyrenees, the Alps, the Apennines, the Carpathians, the Caucasus and the Balkan mountain range (Temple and Terry 2007).

There are 219 species of terrestrial mammals in Europe, of which 59 (26.9%) are endemic (Temple and Terry 2007, Mitchell-Jones et al. 1999). Biogeographic studies indicate that there are low levels of mammalian diversity but high levels of endemicity compared to Asia and North America (Maridet et al. 2007). Europe has 23 species of shrews (Temple and Terry 2007), 12 of which belong to the genus *Sorex* which has 74 extant species worldwide (Rzebik-Kowalska 2005).

The biogeographical patterns that explain mammal diversity and distribution in the European peninsula include: 1) a 'peninsular effect', with highest diversity in east-central regions and a lower diversity westwards and northwards, possibly as a result of historical and environmental factors affecting mammal distributions during the Quaternary; 2) isolation and allopatric speciation during the glaciations in the Pleistocene; 3) species evolution and presence of relict species from the Tertiary; and 4) ancient and current human pressures, introduction and extinction of species (Baquero and Tellería 2001).
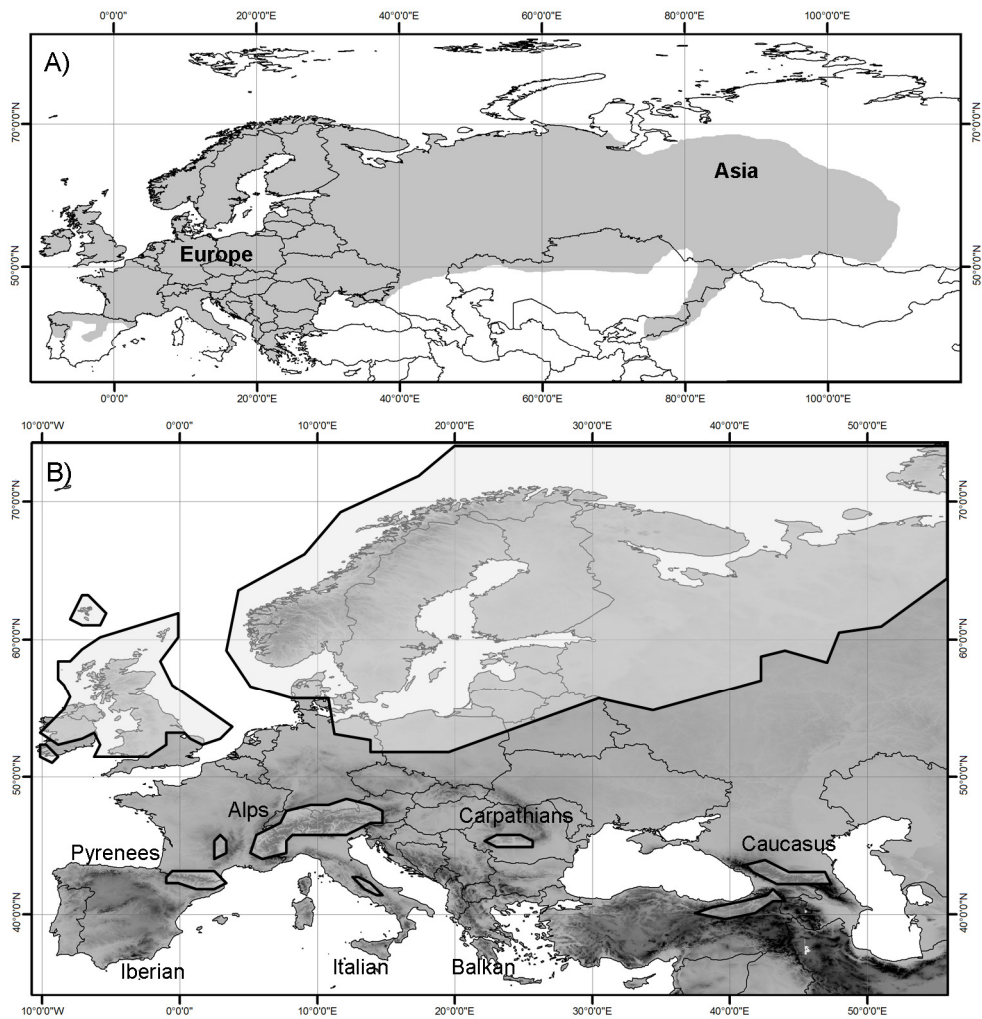
**Figure 1.1.** A) Map of Eurasia showing the geographical distribution of the Eurasian pygmy shrew *Sorex minutus* (modified from Corbet and Harris 1991). B) Map of Europe showing the three Mediterranean peninsulas (namely Iberian, Italian and Balkan) and the glaciated areas at the Last Glacial Maximum (grey shaded polygons delimited by thick black lines; modified from Ehlers and Gibbard 2004).

## 1.3 Shrews including *Sorex minutus*

Shrews are amongst the smallest of all extant mammals, frequently confused with mice because of their small size, short legs, small rounded ears, short dense fur, usually dark colour, and long snouts. However, shrews are classified as insectivores, but do not specialize on insects and may be considered predatory generalists (Crowcroft 1957, Churchfield 1990).

One of the most speciose groups of mammals, shrews occupy moist and terrestrial environments at different altitudes (from sea level to 3000 m.a.s.l), where there is adequate vegetation cover and diversity of invertebrate prey, are found in woodlands, forests, grasslands, hedgerows and scrublands in temperate and tropical regions, and some are aquatic (Churchfield 1990, Corbet and Harris 1991).

Among the primitive characters shown by shrews are the generalized and unspecialized body plan, plantigrade mode of locomotion, each foot with five digits terminating in a simple claw, the skull is long and narrow with a small brain with small cerebral hemispheres but relatively large and well developed olfactory lobes, there are no zygomatic arches, the mandible has a double articulating surface, the genital and urinary systems have a common cloaca, and the testes are retained in an abdominal cavity and do not descend to a scrotum (Churchfield 1990). The teeth are specialized and characteristic with large prominent incisors, which, unlike rodents, do not keep growing all their life span. The first upper incisors are enlarged and have two cusps that project downward, plus a large cusp at the base, and the first lower incisors are also large and slightly procumbent, pointing forward and upwards (Churchfield 1990).

*Sorex minutus*, commonly known as the Eurasian pygmy shrew, has a wide Eurasian range (Fig. 1.1A) and it is the most widespread species of the Holarctic genus *Sorex*, extending from Northern Portugal across most of Central and Northern Europe to Lake Baikal in Siberia. This species is present in the three Mediterranean peninsulas of Iberia, Italy and the Balkans, but is limited to higher altitudes where its distribution becomes patchy (Hutterer 2005). However, humans are also believed to be involved in the distribution of this species, particularly for the populations in Ireland and the islands off the coast of Britain (Mascheretti et al. 2003). The widespread distribution of *S. minutus* makes it an excellent model for studies of genetic and morphological evolution in a continental and island setting.

4

To classify taxonomically the study species, I follow the recent splitting up of the paraphyletic Order 'Insectivora' into the Orders Soricomorpha (moles, shrews and *Solenodon*), Erinaceomorpha (hedgehogs) and Afrosoricida (tenrecs and golden moles) (Stanhope 1998, Hutterer 2005). Then, I follow the two-family classification of shrews by Reumer (1998), which distinguishes the extinct Heterosoricidae and the extant Soricidae. Therefore, the Eurasian pygmy shrew *Sorex minutus* (Linnaeus 1766) belongs to: Class Mammalia, Order Soricomorpha, Family Soricidae, Subfamily Soricinae, Tribe Soricini (with a Holarctic distribution), Genus *Sorex*, Species *minutus*. The morphological features of *S. minutus* described here follow Crowcroft (1957): the adults have a nose to anus length of 55-66 mm, tail length of 30-45 mm, hind foot of 10-11.5 mm and weight around 3-6 g; the juveniles have a length of 45-55 mm and weight between 2.5 and 4.5 g. The tail is hairy and usually exceeds 2/3 of the body length. The colour is sandy brown and the ears are never white, as found in some species (Crowcroft 1957, Corbet and Harris 1991). *Sorex minutus* has 339% of the expected basal metabolic rate for mammals of that body weight (Sparti and Genoud 1989), and it requires 125% of its body weight daily in food or 9.7-13.0 KJ per gram body weight per day (Churchfield 1990). The diet of *S. minutus* includes flies, spiders, woodlice and earthworms, and less frequently millipedes and centipedes (Churchfield 1990).

Five valid subspecies, which should be regarded with caution, were recognized by Hutterer (1990): *S. m. minutus* in Northern and Central Europe to Siberia, *S. m. gymnurus* (Chaworth-Musters 1932) in Greece, *S. m. becki* (von Lehman 1963) in the Alps, *S. m. carpetanus* (Rey 1971) in Spain, and *S. m. lucanius* (Miller 1909) in the Basilicata and Calabria ridges of Southern Italy; however, he did not resolve the taxonomic status of pygmy shrew populations in Central Italy. The populations appear to follow a clinal gradient within regions (Hutterer 1990). The karyotype has a diploid number (2n) of 42 chromosomes, and

a fundamental number of 56, but 2n = 40 and 36 are found in the Baltic islands of Öland and Gotland (Hutterer 2005).

The fossil record of the Soricidae in Europe starts in the Early Oligocene, while the oldest record of the Soricini in Europe dates probably to the Late Miocene, including the genus *Sorex* (Rzebik-Kowalska 1998). The oldest record of *S. minutus* was found in Podlesice and Mała Cave, Poland in the Early Pliocene (Mammal Neogene 14), between 5.3 and 3.6 MYA (Rzebik-Kowalska 1998).

**1.4 The Quaternary glaciations**

For this study, a geological timeframe is needed to fully understand the main findings. Most of the events discussed here occurred within the Quaternary period (2.58 – 0.0117 MYA; Gibbard et al. 2010), commonly recognized as the ice ages. The Quaternary was marked by significant changes in climate that led to periods of ice-sheet expansion or 'glaciations', and intervening temperate (warm) stages or 'interglacials' (Ehlers 1996), with profound effects on the current distribution of European biota and their intraspecific genetic variation (Hewitt 2000, 2004). The Quaternary is further subdivided in two epochs: the Pleistocene (2.58 – 0.0117 MYA) with the subepochs Early, Middle and Late, and the Holocene (11.7 KYA to present) (Gibbard et al. 2010).

About 30 warm and 30 cold stages are distinguished in the Pleistocene, with cold stages every ca. 41 KY cycles from the beginning of the Pleistocene until ca. 0.9 MYA and every 100 KY from then to the present (Hewitt 2000), caused by periodic changes of the Earth's orbit around the Sun that affect the influx of solar radiation. The factors affecting the radiation influx are: 1) the Earth's eccentricity or deviation from a circular orbit, which enhances the contrast in temperature between summer and winters, and varies between 0.5 to 6% within 100,000 years; 2) the tilt of the Earth's axis, which changes over 40,000 years between 22.1° and 24.5°; and

3) the Earth's precession, or rotation of the Earth's axis around the pole, which together with the orbital ellipse determine the perihelium with a periodic change every 20,000 years and enhances the annual temperature gradient on one hemisphere and smoothes it on the other (Ehlers 1996). However, the mechanisms that cause the climatic fluctuations may be more complex involving a combination of astronomical, geological and biological factors.

During the Quaternary ice ages substantial areas of Northern Europe were covered by ice sheets (Fig. 1.1B) 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 recolonised Central and Northern Europe from the traditionally recognized southern glacial refugia in response to the late-glacial and postglacial warming, and this process became established as a biogeographical paradigm (Taberlet et al. 1998, Hewitt 2000, 2004). In addition, there is evidence supporting glacial refugia for some temperate and boreal species further north than traditionally recognized, from where the postglacial colonisation of Central and Northern Europe occurred, implying a more complex pattern of glacial survival and postglacial recolonisation (Bilton et al. 1998, Stewart and Lister 2001, Kotlík et al. 2006, Stewart et al. 2010).

The isolation and recolonisation during the climatic fluctuations of the Quaternary must have had profound effects on population demographic parameters and must have left genetic signatures that can be explored using a phylogeographic approach. Moreover, the latitudinal and altitudinal range shifts and isolation of populations because of the presence of geographical barriers to dispersal such as glaciers could have provided an opportunity for adaptation to occur, which could have affected the intraspecific levels of morphological variation.

**1.5 Phylogeography defined**

The term phylogeography implies the integration of phylogenetic analyses of organismal data with a geographical context. This term was coined by Avise et al. (1987) and stemmed from the idea that species are genetically structured due to environmental, geographical, geological or behavioural factors. Phylogeography is generally defined as "...a field of study concerned with the principles and processes governing the geographic distributions of genealogical lineages, especially those within and among closely related species." (Avise 2000).

Phylogeography is considered as a subdiscipline of biogeography, bridging the gap between microevolutionary and macroevolutionary disciplines, and it is used to infer historical (evolutionary) patterns, gene flow, colonisation and genetic bottlenecks among populations within and between species, and to delimit species boundaries and define conservation units (Domínguez-Domínguez and Vázquez-Domínguez 2009). It is an integrative field of evolutionary biology because it exploits very different scientific disciplines such as molecular biology, population genetics, geographical information systems, geology, taxonomy and systematics, conservation biology, ecology, etc.

Traditionally, the topology of the phylogenetic trees, the geographic distribution of lineages and the estimated dates of branching events of the trees were used to provide a direct qualitative interpretation of the demographic history of the taxa under study (Hickerson et al. 2010). The favoured statistics were based on the hierarchically partitioned Wright's F statistics (Wright 1969) and Nei's summary statistics (Nei and Li 1979, Nei 1987). Since then, the field of phylogeography has gone through a methodological and philosophical transformation in how the genetic data are used to infer the demography. The breakthrough was the shift to the use of genealogy samplers such as the coalescent to derive parameter estimates.

The coalescent, or simply coalescence, is the retrospective process in which genes from the present merge into their common ancestor going back in time, and it is represented as a gene genealogy of a sample of individuals taken from a population (Knowles 2009). The gene genealogy can be modelled using a priori information, or knowledge of the biological system under study, or it can be modelled using a variety of possible demographies. When using different models to infer population demographic history, no competing model is true, but it is instead a probabilistic approximation to a particular demographic history that explains better the observed patterns of the gene genealogies (Hickerson et al. 2010). The use of statistical approaches based on coalescent models that take into account the stochasticity of genetic processes for parameter estimation and hypothesis testing is termed statistical phylogeography (Knowles and Maddison 2002).

The marker of choice for microevolutionary phylogenetic inference since the development of phylogeography has been the mitochondrial (mt) DNA because of its rapid sequence evolution (Avise 2000). However, there has also been a shift towards the use of multilocus data in statistical phylogeography to provide a more robust basis for statistical inference (Knowles 2009), usually incorporating nuclear genes or strictly paternally non-recombining inherited molecular markers, such as Y-chromosome introns (Hellborg and Ellegren 2003). The mitochondrion is an intracellular organelle that has its own genome, and it can move, divide and fuse within a cell. The genome consists of 15,000 – 17,000 bp of a double stranded and circular DNA molecule that encodes 37 genes. Twenty-four genes encode the translational molecules for the mtDNA (22 transfer RNAs and two ribosomal RNAs) and 13 genes encode subunits of the electron transfer chain involved in cellular respiration and ATP production (Ballard and Whitlock 2004). The mtDNA is characterized by the lack of recombination, asexual inheritance and almost strictly maternal mode of inheritance (no heteroplasmy, although there are a few

exceptions), exists in high number of copies per cell and has extensive intraspecific polymorphism, all of which are useful for estimating population demographic parameters (Avise 2000). Within the mtDNA, the cytochrome (cyt) *b* gene has been the marker of choice for phylogeographic studies of recently separated mammalian taxa because of its moderate substitution rate, compared to the hypervariable mtDNA control region, for example, and because its levels of genetic divergence are in the phylogenetically informative range without saturation effects (Johns and Avise 1998). The cyt *b* is the central catalytic subunit of the ubiquinol cytochrome *c* reductase complex and encodes a transmembrane enzymatic protein subunit involved in the electron transfer of the cellular respiratory chain of all eukaryotic organisms (Esposti et al. 1993).

## 1.6 Geometric morphometrics defined

Morphometrics, considered as a subfield of statistics or a branch of mathematical shape analysis, is an old discipline etymologically defined as the measurement of shape (Mitteroecker and Gunz 2009). The objective of morphometrics is to characterize, analyze and compare biological forms quantitatively. Traditionally, the measurements that needed to be taken from an object to describe its shape were the length, depth and width, but more recently the field of morphometrics experienced a 'revolution' by the invention of coordinate-based methods, the development of the statistical theory of shape and the use of transformation grids to visualize shape variation (Rohlf and Marcus 1993, Adams et al. 2004, Mitteroecker and Gunz 2009). This new morphometrics approach was termed 'geometric morphometrics' because it preserves the geometry of the landmark coordinates throughout the analysis. Geometric morphometric methods give a precise and accurate description of shape variation, and a better

visualization, interpretation and communication of the results in comparison to traditional morphometrics (Zelditch et al. 2004).

Under the statistical theory of shape, any object has a form that is composed of shape and size. Shape is mathematically defined as all the information that is independent of the object's size, position and orientation, while size is the scale of the object, and it can be measured by any interlandmark distance or by a series of single size measures (Mitteroecker and Gunz 2009). Size measurements and the analysis of shape variation depend on the selection of landmarks, which are defined as discrete anatomical points present in all the specimens in the study, i.e., biological homologous anatomical loci that must be represented on each object and must exist between and within populations (Dryden and Mardia 1998). Landmarks contain the information about their relative position in their Cartesian coordinates and simple algebraic manipulations can separate that information into independent statistically uncorrelated components of size and shape; that information can then be used as variables for statistical analysis (Zelditch et al. 2004).

## 1.7 Aims

In this study, two different techniques and areas of scientific research are to be implemented for the understanding of the genetic and morphological diversity of *S. minutus*. First, a phylogeographical approach will help understand several aspects of the past and current distribution of the genetic lineages. Second, a geometric morphometrics approach will examine the morphological variation within and among the genetic lineages. The use of both approaches can give very powerful insights regarding the evolutionary history of this small mammal. The knowledge acquired through this project will be a significant piece of the puzzle for the understanding of the recent evolutionary history of European small mammals.

The general aims of this study are:

- To estimate the phylogeographic structure (i.e., the geographical distribution of genealogical lineages) of the Eurasian pygmy shrew *Sorex minutus* using different molecular markers. Where different lineages are found, to estimate the genetic diversity and structure, time of divergence and historical demographic processes of those lineages.

- To determine the glacial refugia and colonisation routes for *S. minutus* in Europe using a more detailed study than previously, and compare the results with the phylogeographic paradigms of glacial refugia in the Southern Mediterranean peninsulas.

- To estimate the morphological variability *of S. minutus* in Europe using a geometric morphometrics approach and compare it with the phylogeographic structure. Where significant differences are found, to identify any geographical or genetic pattern to which it could be related.

- To explore the origin, colonisation routes, genetic diversity and structure, and morphological variation of *S. minutus* in the British Isles.

# Chapter 2

# Phylogeography of *Sorex minutus* in Europe

## 2.1 Introduction

It is widely recognized that the three Southern European peninsulas, namely the Iberian, Italian and Balkan (Fig. 1.1B), acted as Mediterranean glacial refugia for many temperate and warm-adapted species during the Quaternary ice ages (Hewitt 2000). Under this scenario, populations and species in the southern peninsulas remained isolated during cold stages by ice sheets covering most of the Alps and the Pyrenees and by unsuitable climatic conditions further north, which acted as geographical barriers to dispersal. Different genealogical lineages evolved in these refugial areas; thus, the Southern European peninsulas became current hotspots of intra-specific diversity (Bilton et al. 1998, Hewitt 2000, Myers et al. 2000, Petit et al. 2003). As glaciers retreated and suitable habitat became available, it is also posited that populations at the northern limits of the southern refugia expanded and colonised territories further north, but with subsequent loss of alleles and heterozygosity because of founder events at the leading edge, whereas populations in the south maintained higher genetic diversity (Hewitt 1996).

The concept of glacial refugia has been mostly applied to the isolation and differentiation of species and/or populations in Mediterranean peninsulas during cold phases of the Quaternary, with glaciations as the major force shaping population divergence in allopatry (Stewart et al. 2010). However, it has been shown that this biogeographic pattern is further complicated by the existence of cryptic northern glacial refugia, further north than the traditionally recognized southern refugia. There is palaeontological, palynological, phylogeographic and species distribution modelling (SDM) evidence for some temperate and boreal species that support this hypothesis of 'northern refugia' (Bilton et al. 1998, Willis

and van Andel 2004, Kotlík et al. 2006, Magri et al. 2006, Sommer and Nadachowski 2006, Svenning et al. 2008, Fløjgaard et al. 2009, Vega et al. 2010b). Moreover, recent findings of intra-specific genetic and morphological differentiation within multiple micro-refugia in Iberia (Gómez and Lunt, 2007), Italy (Canestrelli et al. 2007, Canestrelli et al. 2008) and the Balkans (Kryštufek et al. 2007) add to the current understanding of the genetic diversification and biological diversity in southern refugia. Additionally, interglacial refugia, the reverse phenomenon of glacial refugia, have been hypothesised to explain the persistence of cold adapted species through interglacials (Hilbert et al. 2007). Hence, the term Quaternary refugia can be broadly defined as the geographical region that a species inhabits during periods of adverse climate (either glacial or interglacial) that represent the species' maximum contraction in geographical range (Stewart et al. 2010), highlighting the idea that different species can respond to climate change independently.

Previous phylogeographic studies on *Sorex minutus* using mitochondrial (mt) DNA and Y-chromosome introns revealed a genetic structure over Eurasia, with isolation and differentiation in different southern and northern refugia (Bilton et al. 1998, Mascheretti et al. 2003, McDevitt et al. 2010). Bilton et al. (1998) analysed a partial sequence of the cytochrome (cyt) *b* gene (581 bp) based on 14 individuals and found 13 haplotypes distributed in four clades: three geographically restricted lineages corresponding to 'Balkan', 'Central Italian' and 'Iberian' clades, and a 'major lineage' linking together individuals from Northern-Central Europe to Lake Baikal in Siberia, which gave support to the hypothesis of northern refugia. However, these authors used a restricted number of samples and did not cover the full extent of the geographical range of the species. In a more detailed study, Mascheretti et al. (2003) explored the colonisation of Ireland based on 74 specimens of *S. minutus* with 1110 bp of the cyt *b* gene. They recovered a similar

tree topology as previously, with a northern and three peninsular lineages, but also an 'Andorran-Irish' lineage, interpreted as evidence for human-mediated colonisation of Ireland from South-Western Europe. The sequence similarity between Ireland and the Pyrenees resembled a Lusitanian element, linking together several species from South-Western Ireland with Portugal and North-Western Spain (Corbet 1961, Yalden 1982). McDevitt et al. (2009, 2010) analysed a larger dataset of *S. minutus* using microsatellite, mitochondrial and Y-chromosomal molecular markers and found that the Irish populations were introduced most likely in the Neolithic period, had smaller genetic diversity than the British population and presented a signature of rapid population expansion in Ireland. They also found a similar tree topology with the mitochondrial Control Region as found previously with cyt *b*, showing Northern, Balkan, Iberian, Italian and Western (composed of Andorran-Irish-French) lineages, plus a single individual with distinct mitochondrial and Y-chromosome DNA from Southern Italy. Additionally, Vega et al. (2010a) showed that the South Italian pygmy shrew is genetically and morphologically different from populations in Central and Northern Italy and Europe, and Vega et al. (2010b) showed that the northern lineage has a strong signature of population expansion and it is genetically different from southern lineages, favouring the northern refugia hypothesis. However, more sampling in southern peninsulas and Northern-Central Europe is needed to rule out completely the hypothesis of southern peninsulas as source areas for the northern lineage, to estimate the amount of differentiation, colonisation routes and contact zones among them after the Last Glacial Maximum (LGM), to pinpoint the origin of Irish and British populations and to evaluate the demographic parameters within peninsulas and other lineages.

Here, I undertake a statistical phylogeographic approach to further this aim. In particular, I address the following specific study questions: 1) Does a more

detailed phylogeographic analysis, using more samples and molecular markers, detect the same genetic lineages that have been previously found? 2) Do different molecular markers recover the same phylogenetic tree topology or distinguish the same genetic lineages? 3) What is the geographic distribution of the genetic lineages of *S. minutus* within Europe, and are there areas of contact between the different lineages? 4) What is the genetic diversity within the lineages and are they significantly structured over Eurasia? 5) Is the South Italian pygmy shrew genetically distinctive from other Italian populations, as suggested by previous analyses? 6) What were the colonisation routes, times of divergence and historical demographies of the genetic lineages? 7) Do the genetic lineages show a signature of population expansion or have the populations remained constant through time? 8) Are the new data on the northern lineage still compatible with the northern refugia hypothesis? 9) Can the origin of the Irish and British populations be determined with more detailed sampling in continental Europe? To answer these questions I employ a substantial battery of software for the analysis of genetic variation and three molecular markers, the cyt *b* gene, two Y-chromosome introns and the nuclear BRCA1 gene, on a larger number of samples than previously.

The detailed study of phylogeography of *S. minutus* has broader significance. The nature of southern and northern glacial refugia has important implications for the understanding of their biogeographical roles as sources of genetic diversity, areas of speciation, identification of conservation units and preservation of species, and the results presented here should be put in the wider context of European biodiversity. Such information is relevant for understanding how species responded to past climate changes and may be useful to forecast how climate change may affect the current distribution of species, and contributes towards an emerging new synthesis of the full-glacial distributions of the European biota.

**2.2 Materials and methods**

2.2.1 Samples and GenBank sequences for molecular analysis

A total of 544 samples and sequences were obtained for the cyt *b* analysis (Appendix 3). Of that total, 236 *S. minutus* samples, plus two *S. volnuchini* samples used as an outgroup, were obtained from fieldwork, museum collections or were kindly provided by colleagues (this includes 78 *S. minutus* samples distributed among 68 haplotypes deposited in GenBank: GQ494305 – GQ494350, GQ272492 – GQ272518, Vega et al. 2010a, b). Three hundred and five *S. minutus* cyt *b* sequences were obtained from studies reported by other researchers, using haplotypes deposited in GenBank (AJ535393 – AJ535457, Mascheretti et al. 2003; AB175132 and AB175133, Ohdachi et al. 2006; FJ623774 – FJ623893, Searle et al. 2009), plus one cyt *b* sequence of *S. volnuchini* (AJ535458, Mascheretti et al. 2003).

A total of 127 sequences of DBY-7 and UTY-11 introns were obtained for the Y-chromosome analysis (Appendix 3). From the total, 71 DBY-7 and UTY-11 sequences were obtained from *S. minutus* samples from fieldwork, museum collections and provided by colleagues. This includes 11 sequences reported in Vega et al. (2010a) distributed among the following eight sequence variants deposited in GenBank (GQ272519 – GQ272521 for DBY-7 and GQ272522 – GQ272526 for UTY-11). Data on 55 sequences, plus one sequence of *S. volnuchini*, were obtained from McDevitt et al. (2010), using the following sequence variants recorded in GenBank: EU564415 – EU564422 for DBY-7 and EU564423 – EU564434 for UTY-11.

For the BRCA1 gene, 36 new sequences were obtained (Appendix 3), plus one sequence of *S. volnuchini* from GenBank which was used as outgroup (DQ630236, Dubey et al. 2007).

Fieldwork was done in several regions of Southern, Central and North-Western Italy, Northern Spain and Ireland to improve the sample size of previous studies and to cover sampling gaps in Europe. Tissue samples from the specimens collected were placed in 90% or 100% Ethanol until use, and whole body specimens were deposited at the National Museums of Scotland for long term storage. Tissue samples from museum specimens were obtained by toe clipping or by cutting a small piece of skin, to avoid damage to the specimen, or by removing residual desiccated brain tissue from skulls using a surgical curette.

## 2.2.2 DNA extraction, gel electrophoresis and visualization

*Sorex minutus* DNA was obtained from fresh and collection samples using the Blood and Tissue DNA Extraction Kit (Qiagen) following the protocols provided and stored at –20 °C. Prior to extraction, fresh ethanol-preserved samples were dried at room temperature and digested at 56 °C for ca. 5 h in the cell lysis solution with 20 mg/ml Proteinase K provided with the kit, while museum samples that might have been stored in Formalin were dried first and then digested overnight. To visualise the quality of the extracted DNA, 5 µl of sample mixed with 2 µl of Loading Buffer were run in a 1% Agarose 1X TAE (Tris base-Acetic acid-Ethylenediaminetetraacetic acid pH 8.0) buffer gel at 100 V for 35 min, next to a standard ladder containing known molecular weights for comparison. The gel was placed under UV light and photographed.

Good quality DNA, usually from fresh samples or well preserved specimens, appears in the Agarose gels as a single bright band of high molecular weight, evidence of no degradation and good extraction yields. Degraded DNA, on the other hand, appears as a smear of varying molecular weights. Old museum samples usually appear as a light smear of low molecular weight between 600 and 200 bp.

2.2.3 PCR amplification, purification and sequencing

Partial cyt *b* sequences from good quality DNA (obtained from fresh samples or well-preserved museum specimens) were obtained by the Polymerase Chain Reaction (PCR) using two primer (oligonucleotide) pairs that amplified ca. 700 bp of overlapping fragments. Partial cyt *b* sequences from samples with highly degraded DNA were obtained by PCR using a combination of five primer pairs that amplified an overlapping set of fragments of ca. 250 bp (Table 2.1).

**Table 2.1.** Primer sequences used in this study.

| Sequence | Forward primer 5'-3' | Reverse primer 5'-3' | Approx. fragment length (bp) |
|---|---|---|---|
| Cyt *b* | GACAGGAAAAATCATCGTTG | GTAGTAGGGGTGGAAAGGAATT | 700 |
| Cyt *b* | GAGGACAAATGTCATTCTGAGGC | TTCATTACTGGTTTACAAGAC | 700 |
| Cyt *b* | GACAGGAAAAATCATCGTTG | CGACGTGAAGGAATAGACAGA | 250 |
| Cyt *b* | CCGAGACGTAAACTACGGATG | AGAAGCCCCCTCAGATTCAT | 250 |
| Cyt *b* | TTCTGAGGCGCAACAGTTATT | AATGAGGAGGAGCACTCCTAGA | 250 |
| Cyt *b* | CCTGTCCTCAGACGCAGATA | GTTTGGAGGTGTGGAGGAAA | 250 |
| Cyt *b* | TCCCGAATAAACTAGGAGGTGT | TTCATTACTGGTTTACAAGAC | 250 |
| DBY-7[1] | GGTCCAGGAGARGCTTTGAA | CAGCCAATTCTCTTGTTGGG | 500 |
| UTY-11[1] | CATCAATTTTGTAYMAATCCAAAA | TGGTAGAGAAAAGTCCAAGA | 1000 |
| BRCA1[2] | TGAGAACAGCACTTTATTACTCAC | ATTCTAGTTCCATATTGCTTATACTG | 750 |

[1]Hellborg and Ellegren (2003).

[2]Dubey et al. (2006).

PCR amplification of cyt *b* was performed in a 50 µl final volume containing: 1X Buffer, 1 µM each primer, 1 µM dNTP's, 3 mM $MgCl_2$ 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, using a C1000 thermal cycler (BIO-RAD). Negative controls were included during the PCR amplification to check for contamination.

PCR amplification of the Y-chromosome introns was done using one primer pair for each intron, DBY-7f with DBY-7r and UTY-11f with UTY11r (Table 2.1,

Hellborg and Ellegren 2003), in a 25 μl final volume containing: 1X Buffer, 5 μM

each primer, 2 μM dNTP's, 2.5 mM MgCl$_2$ and 0.25 U Platinum *Taq* Polymerase

(Invitrogen), with cycling conditions: 95 °C for 10 min, 20 cycles of 95 °C for 30 s,

touchdown temperatures from 65 – 55 °C (for UTY-11) or 55 – 45 °C (for DBY-7) for

60 s decreasing by 0.5 °C/cycle and 72 °C for 90 s, followed by 20 cycles of 95 °C

for 30 s, 55 °C (for UTY-11) or 45 °C (for DBY-7) for 60 s and 72 °C for 90 s, and a

final elongation step at 72 °C for 10 min (Hellborg and Ellegren 2003). Known

female and male samples were used as controls, and negative controls were

included in all PCRs for the Y-chromosome introns to check for contamination. The

sequences of DBY-7 and UTY-11 introns were concatenated and analysed jointly.

Amplification of the BRCA1 nuclear gene was performed using the primer

pair BRCA1f and BRCA1r (Table 2.1, Dubey et al. 2006) in a 50 μl final volume

containing: 1X Buffer, 0.6 μM each primer, 1 μM dNTP's, 3 mM MgCl$_2$ and 0.5 U

Platinum *Taq* Polymerase (Invitrogen), with cycling conditions: 94 °C for 4 min, 40

cycles at 94 °C for 60 s, 50 °C for 60 s and 72 °C for 120 s, and a final elongation

step at 72 °C for 7 min (Dubey et al. 2006). Negative controls were included in all

PCRs to check for contamination.

The purification of PCR products was done with a commercial kit (Qiagen).

DNA was quantified by spectrophotometry in a NanoDrop 1000 (Thermo Scientific)

at 260 nm and sequenced in an ABI 3730XL DNA Sequencer using Big Dye

Terminator chemistry (Macrogen, Inc. and Cornell University Core Laboratories

Center).


2.2.4 Analysis of sequence polymorphism

DNA sequences of cyt *b*, Y-chromosome introns and BRCA1 were edited in

BioEdit version 7.0.9.0 (Hall 1999) and aligned using ClustalX version 2.0 (Larkin et

al. 2007). The software DnaSP version 5.10 (Librado and Rozas 2009) was used to

calculate for all molecular markers: the number of monomorphic and polymorphic sites, sites with alignment gaps, the number of parsimony informative sites and singleton mutations, and the number of haplotypes or sequence variants within the sample; and for the cyt *b* data only: the haplotype frequency and the number of synonymous and non-synonymous substitutions in the codons using a conservative approach. The parsimony informative sites were counted as the number of sites where at least two types of nucleotides were found and were represented at least twice in the sample. The singleton mutations were counted as the number of sites where only one mutation was found in all the samples.

The neutral hypothesis for the cyt *b* gene was tested between *S. minutus* and the outgroup using the McDonald-Kreitman (MK) test (McDonald and Kreitman 1991), where the ratio of non-synonymous to synonymous fixed substitutions among populations should be the same as the ratio of non-synonymous to synonymous polymorphisms within populations if evolution of the protein coding gene occurs by neutral processes.

2.2.5 Phylogenetic analyses

The haplotypes and sequence variants found in cyt *b*, Y-chromosome introns and the BRCA1 gene datasets were used for the phylogenetic reconstructions. A phylogeny is a graphic representation of the evolutionary relationships among species or, in this case, among individuals of the same species.

The use of models of evolution, or models of nucleotide substitution, allows for different probabilities of change among nucleotides, and the failure to allow unequal substitution rates can lead to underestimates of the genetic distances between sequences and problems with long branch attraction (Waddell and Steel 1997). To identify the most appropriate substitution model for the molecular data

available, I used the software jModelTest version 0.1.1 (Posada 2008) that statistically selects the best-fit model with a hierarchical likelihood ratio test. I set a fixed tree with the BIONJ algorithm (a variation of the Neighbor-Joining method described below; Gascuel 1997) to estimate the branch lengths as model parameters, and used the Bayesian Information Criterion (BIC) value to select the best of 56 candidate models. The BIC selects the model with the maximum posterior probability when all models have the same prior.

The model of evolution that best fit the cyt $b$ data was the General Time Reversible (GTR) with a gamma shape parameter ($\Gamma$) of 0.2720 and no specified proportion of invariable sites (I), base frequencies for A = 0.2819, C = 0.3062, T = 0.1318 and G = 0.2800, and six substitution types (A–C = 0.2384, A–G = 14.4125, A–T = 0.7193, C–G = 1.3882, C–T = 9.2882, G–T = 1.000). Similarly, the Akaike Information Criteria, with a weaker penalization for increased model complexity, suggested the GTR + $\Gamma$ + I. The GTR model is a distance calculation method that gives the expected average number of substitutions per site by comparing all pairs of sequences (Lanave et al. 1984, Waddell and Steel 1997). It allows unequal substitution rates across sites because of invariable sites, incapable of changing due to biological constraints (Waddell and Steel 1997), and accommodates rate heterogeneity among codon positions by fitting a gamma distribution (i.e., a probability model) with an alpha shape parameter which is inversely correlated with increasing rate heterogeneity (Xia 1998).

For the concatenated Y-chromosome introns, the best fit model was the Hasegawa-Kishino-Yano (HKY, Hasegawa et al. 1985) with no specified proportion of invariable sites, no gamma correction, equal rates for all sites, a transition/transversion ratio = 1.6681 and nucleotide frequencies of A = 0.2832, C = 0.1904, G = 0.1969 and T = 0.3295. For the BRCA1 gene, the model selected was again the HKY, with no gamma correction and equal rates for all sites, a

transition/transversion ratio = 12.1998 and nucleotide frequencies of A = 0.3825, C = 0.1749, G = 0.2130 and T = 0.2296. The HKY model allows for a different transition/transversion ratio and unequal base frequencies, and is a generalization of simpler models of evolution (Felsenstein 2004).

The phylogenetic relationships among cyt *b* haplotypes and among nuclear markers sequence variants were inferred with a variety of traditional phylogenetic methods, including Neighbor-Joining (NJ), Maximum Parsimony (MP) and Maximum Likelihood (ML). I used PAUP* version 4.0b10 (Swofford 2000) complemented with PAUPUp version 1.0.3.1 (by F Calendini and J-F Martin, available at http://www.agro-montpellier.fr/sppe/Recherche/JFM/PaupUp/main.htm) to infer NJ trees with the BIONJ method (described below; Gascuel 1997) and the appropriate evolutionary model, and MP trees with a heuristic search with simple stepwise addition of taxa, branch swapping (tree bisection reconnection, one million rearrangements) and 100 random replicates. ML trees were inferred with RAxML version 7.0.4 software (available at http://www.phylo.org/news/RAxML; Stamatakis 2006) using the appropriate evolutionary model.

In brief, NJ is a fast distance clustering method that does not assume a molecular clock and it is based on the principle of minimum evolution, where the sum of the branch lengths is minimized (Felsenstein 2004). NJ uses an agglomerative algorithm that chooses two sequences iteratively, creates a new node that represents the cluster of the two sequences and reduces the distance matrix by replacing the two sequences by this node, and repeats the operation until all sequences are chosen. The BIONJ method is a variant of the NJ method where the variances and co-variances of the evolutionary distances of the distance matrices are reduced (Gascuel 1997). MP reconstructs a phylogeny with the fewest evolutionary events possible, and searches among all possible phylogenies for the one that minimizes the number of evolutionary events making use of the parsimony

informative sites (Felsenstein 2004). This method does not take into account the substitution rate among nucleotides and it uses parsimony as its evolutionary model. ML is a method for the statistical estimation of phylogenies that evaluates different hypotheses of evolutionary history (i.e., phylogenies or trees) in terms of probability of the observed data given a certain evolution model, taking into account the substitution rates among nucleotides, where the phylogeny with the highest probability (or maximum likelihood) of reaching the observed data is preferred over the others (Felsenstein 2004).

Statistical support for the phylogenetic relationships was assessed by 10,000 bootstrap replicates for NJ and 1000 bootstrap replicates for MP and ML. The bootstrap involves resampling from the nucleotide site data with replacement to assess the uncertainty of the estimated phylogeny. With this method, sites are drawn independently from the aligned full dataset, replaced and resampled until a new dummy dataset with the same number of sites is created from which a new tree can be built, and the process is repeated several times. Then, the topology (i.e., the branching pattern) of the generated trees is compared and a proportion value is given to all the branches (Felsenstein 2004); if a branching pattern is found several times in the bootstrap replicates, that branch will have a high proportional value and higher certainty.

Additionally, a Bayesian analysis (BA) for phylogenies was performed in MrBayes version 3.1 (Huelsenbeck and Ronquist 2001) using the appropriate evolutionary model. BA is similar to likelihood methods, but makes use of prior information of the quantities being inferred and implements a Markov Chain Monte Carlo (MCMC) method using the Metropolis-Hastings algorithm (Metropolis et al. 1953, Hastings 1970). The MCMC is run for several thousand generations (or cycles) to approximate the highest probability of the data, where the first tree is stochastically perturbed and the new tree is either accepted or rejected based on

the ratio of probabilities of the new and previous trees (Huelsenbeck et al. 2001).

The acceptance ratio between trees is calculated using the prior probabilities of the new and previous trees, multiplied by the likelihood ratio (Felsenstein 2004). Then, after several generations, the acceptance ratio reaches an equilibrium distribution where the probability of accepting a new tree is smaller than the probability of accepting the previous tree (Felsenstein 2004). All the proposed trees before this equilibrium can be discarded as a "burn-in" and trees can be recorded every several steps during the equilibrium phase. The MCMC in BA draws a random sample from the posterior distribution of trees after burn-in and, with enough samples, it is possible to give a probability value to the branches within a tree (Felsenstein 2004), named as posterior probability.

For the BA of cyt *b* data, two runs were performed with 10 million generations, a sampling frequency every 1000 generations (to give a total of 10,000 samples for each run), a temperature of 0.01 for four heated chains, six rate parameters (equivalent to the GTR model) and checked for convergence with the program Tracer version 1.4.1 (Drummond and Rambaut 2007, available at http://beast.bio.ed.ac.uk/Tracer). Trees were summarised after a burn-in value of 2500 to obtain the posterior probabilities of each phylogenetic split with a 50% majority rule consensus. For nuclear genes and Y-chromosome introns I used similar settings, but changed the temperature to 0.1 and the number of rate parameters to two (equivalent to the HKY model).

To complement the tree-generating programs, a phylogenetic network was constructed only for the cyt *b* data using the software Network version 4.5.1.0 (Fluxus-Engineering) with a Median-Joining (MJ) algorithm and a 'greedy' method (Foulds et al. 1979, Bandelt et al. 1999). The phylogenetic MJ network first creates a minimum spanning network, which is a network with the least number of steps, without reticulations (loops or cycles) among sequences and without inferring

additional sequences, so that the total distance is kept minimal or parsimonious. Then, it adds triplets of closely related sequences connecting them to the minimum spanning network by median vectors (hypothetical nodes) in a 'greedy' manner, where the least number of median vectors is added only if they result in branching nodes (Bandelt et al. 1999).

Phylogenetic networks provide an explicit graphic representation of the evolutionary history between sequences in which taxa are represented as nodes and their evolutionary relationships are represented by edges. The most internal nodes represent ancestral states from which more recent and peripheral nodes derive (Avise 2000). Also, networks have the advantages of representing non-hierarchic evolutionary relationships as multifurcations and loops, being able to work with very few mutations, extant ancestral nodes among haplotypes and large sample sizes, typical characteristics of intraspecific data (Posada and Crandall 2001).

2.2.6 Analysis of sequence diversity

Nucleotide diversity for all sequence sets (cyt *b*, Y-chromosome introns and BRCA1 gene) and haplotype diversity for the cyt *b* dataset only were calculated using DnaSP. Nucleotide diversity ($\pi$) was calculated as the average number of nucleotide substitutions per site between two sequences (Lynch and Crease 1990), but using the Jukes-Cantor correction (Jukes and Cantor 1969) performed in each pairwise comparison. Haplotype (gene) diversity (Hd) was calculated as the frequency of the haplotype in the population, but using the total number of samples, not twice as much (Nei 1987, equation 8.5) because cyt *b* is a haploid (mitochondrial) gene.

2.2.7 Divergence and genetic structure of populations

The amount of divergence and genetic structure was estimated for the phylogenetically coherent groups detected by the cyt *b* phylogenetic analysis and for the geographically delimited group of samples. These groups included the full monophyletic dataset (excluding the outgroup), the phylogroups detected by the phylogenetic methods (including the outgroup as another phylogroup), and continental and island samples. I decided to analyse the samples by phylogroups because I cannot delimit "true" populations based solely on the geographical locations from where the samples were obtained, at least in continental Europe where the distribution of *S. minutus* is continuous. In contrast, island and full continental samples of the phylogroups can be easily distinguished and were analysed separately.

The average number of nucleotide substitutions (Dxy) between two populations (x and y) was calculated using the average number of nucleotide substitutions among all pairs of haplotypes from each population (Nei 1987, equations 10.19 and 10.20). Dxy was used to calculate Da, the net number of nucleotide substitutions between two populations (Nei 1987, equation 10.21). The matrix of Da values, which represent the proportional sequence divergence among populations, was used to construct a NJ tree to depict the evolutionary distances between phylogroups. Da was corrected subtracting the average nucleotide diversity of each population from Dxy, being the corrected Da = Dxy − (Dx + Dy) / 2 (Nei 1987, equation 10.23), where Dx and Dy are the average nucleotide diversity within each population ($\pi$).

The time of divergence between the phylogroups (or groups of samples) was calculated as T = Da / 2$\mu$ (Nei 1987, equation 10.22), where $\mu$ is the nucleotide substitution rate per site per year (using the standard mammal substitution rate of 2% per million years averaged over the whole mtDNA molecule, Wilson et al. 1985)

and Da is the corrected net number of nucleotide substitutions between two populations. The corrected Da values were used because the average time since divergence between haplotypes from two populations is longer than the time of population split, and for recently diverged populations there is a risk of overestimating the divergence time (Edwards 1997). The variance of Da was obtained with DnaSP and used to calculate the standard deviation (SD) for the estimated values of time of divergence between populations.

The geographic midpoints (mid-geographic location between two or more coordinates) for the phylogroups using the samples' coordinates were calculated with the Geographic Midpoint Calculator (available at http://www.geomidpoint.com/). The geographic midpoints were used to obtain the pairwise geographic distances among phylogroups with the Geographic Distance Matrix Calculator version 1.2.3 (by PJ Ersts, available at

http://biodiversityinformatics.amnh.org/open_source/gdmg). A Mantel test was performed in Arlequin version 3.1 (Excoffier et al. 2005) to test the significance of the correlation between the pairwise Da or Wright's $F_{ST}$ genetic structure values (Wright 1969) with the pairwise geographic distances among phylogroups. The significance was obtained through a permutation procedure used to create an empirical null distribution (100,000 permutations).

The genetic structure of populations, including the global and phylogroup specific genetic differentiation ($F_{ST}$) values, pairwise $F_{ST}$ among phylogroups and Analysis of Molecular Variance (AMOVA), was estimated using the software Arlequin. Phylogroup specific $F_{ST}$ indices were calculated as a weighted value of the global population differentiation, which represent the degree of evolution from the ancestral population and are provided as a measure of the contribution of each phylogroup to the global $F_{ST}$. Pairwise $F_{ST}$ among phylogroups, given as a matrix, were calculated as a measure of the degree of genetic differentiation or distance

(Slatkin 1995). The significance values of genetic differentiation were calculated using a non-parametric approach (Excoffier et al. 1992) obtaining a null distribution while permuting haplotypes between populations (10,000 permutations), with $p$ values showing the proportion of permutations giving $F_{ST}$ values larger or equal than the observed ones. The AMOVA follows a similar implementation of the analysis of variance in traditional statistics, but transformed into a hierarchical analysis of molecular data (Weir and Cockerham 1984). The total variance was partitioned into covariance components within the global sample and among phylogroups, and used to compute fixation indices in terms of coalescent times (Slatkin 1995). The significance of the AMOVA was calculated using a non-parametric approach (Excoffier et al. 1992) obtaining a null distribution while permuting haplotypes among phylogroups (10,000 permutations).

The nuclear marker data (Y-chromosome introns and the BRCA1 gene) were not used to estimate genetic divergence and structure among populations because of the low number of informative sites and lack of variation in comparison to the mitochondrial data.

2.2.8 Estimation of historical demographic parameters

Mismatch distributions were obtained with DnaSP for the full cyt *b* dataset and for the main phylogroups computing the distribution of the number of observed pairwise differences among haplotypes and plotting it against their frequency. The mismatch distribution travels as a wave through the horizontal axis, with its crest at an estimated value of $\tau = 2uT$ , where the value T corresponds to the time of population expansion in ½ u generations and u is the mutation rate per sequence (Rogers and Harpending 1992, Rogers 1995). Moreover, the mismatch distribution is unimodal for populations that have been under a sudden expansion, but multimodal for populations with constant population size or demographic equilibrium

(Rogers and Harpending 1992, Slatkin and Hudson 1991). I calculated with DnaSP the theoretical mismatch distribution for a population expansion that would best fit the observed data by implementing a nonlinear least squared approach assuming no recombination, an infinite-sites model (only useful for short periods of time) and that the population size after the expansion is infinite (only reasonable if the population expansion is large) (Rogers and Harpending 1992).

I performed four different tests to verify the hypothesis of sudden population expansion for the full dataset, phylogroups, and continental and island samples. The first method, Tajima's D statistic, tests the null hypothesis that the average number of pairwise nucleotide differences and number of segregating sites in the sample are equal (Tajima 1989). This method explicitly tests for selective neutrality and it is based on an infinite-sites model assuming no recombination; however, significant values can also be related to factors other than selection such as bottlenecks, population expansions or heterogeneity of mutation rates (Tajima 1989). The second method, Fu's Fs statistic, is also a selective neutrality test based on the infinite-sites model, assumes no recombination and tests the probability of having no fewer haplotypes than the observed number in the sample (Fu 1997). This statistic tends to be negative if there is an excess of recent mutations and small pairwise differences among haplotypes (Fu 1997). The third test, Ramos-Onsins and Rozas's $R_2$ statistic (Ramos-Onsins and Rozas 2002), is based on the difference between the number of singleton mutations and the average number of nucleotide differences, where lower $R_2$ values are expected under a population expansion scenario. Coalescent computer simulations were performed in DnaSP to provide confidence intervals and significant values for Tajima's D, Fu's Fs and $R_2$ statistics based on parametric bootstrapping (10,000 replicates). In the fourth method, the occurrence of population expansion was tested using the sum of square deviations between the observed and expected mismatch distributions under

a sudden population expansion model using parametric bootstrapping (10,000 replicates), as implemented in Arlequin. However, Fu's Fs and $R_2$ have been shown to be the most powerful statistics for detecting population growth (Ramos-Onsins and Rozas 2002).

Using the mismatch distributions, the parameters of a sudden population expansion and values of $\tau$ were calculated for the full dataset and all phylogroups with a unimodal distribution. The time of expansion T for the full dataset and phylogroups with a unimodal mismatch distribution was calculated by estimating T from the equation above, being $T = \tau / 2u$ , where $u = 2\mu k$ , $\mu$ is the mutation rate per nucleotide (assuming a mutation rate per nucleotide of 2% per million years, Wilson et al. 1985) per generation (considering one generation per year for *S. minutus*) and k is the number of nucleotide sites (1110 bp for cyt *b* data) (Rogers 1995). It has been shown that this method underestimates $\tau$ and the age of the population expansion because it does not consider mutation rate heterogeneity and uses an unrealistic infinite-sites model (Schneider and Excoffier 1999). Moreover, the assumption of a molecular clock is often invalid, and the use of a phylogenetic rate (an estimate of evolutionary rate calculated by comparing interspecific molecular sequence data) instead of a pedigree rate for intraspecific studies and the time dependency of molecular rates (exponential decay of substitution rate with time) may result in an overestimation of divergence times and population expansion events (Ho et al. 2005, Ho and Larson 2006, but see also Bandelt 2008). Therefore, I also estimated the values of $\tau$ and parameters of population expansion using a generalized non-linear least-squares approach with Arlequin, which takes into account possible heterogeneity of mutation rates and obtains confidence intervals using a parametric bootstrap approach (10,000 replicates) (Schneider and Excoffier 1999).

The use of Bayesian analysis with a relaxed clock model that does not assume a molecular clock and allows for rate heterogeneity can estimate divergence events more accurately. I calculated the Time of divergence from the Most Recent Common Ancestor (TMRCA) for the full cyt *b* dataset using BEAST version 1.5.2 (Drummond and Rambaut 2007) which does a BA using an MCMC algorithm (the input file was created in BEAUti version 1.5.2 included with BEAST). Several short runs were conducted to check the behaviour of the program with the *S. minutus* dataset. Parameters for the final MCMC run reported here were as follows: one run of 50 million generations sampling every 1000 generations (50,000 samples), a prior assumption that the population has remained constant, a random starting tree, a monophyletic group including all *S. minutus* data (without the outgroup), a fossil calibration of 5 MYA (SD = 0.2) with a normal prior distribution, an uncorrelated relaxed clock model (Drummond et al. 2006) to calculate the evolutionary rate, and an GTR substitution model. The trace files were analysed in Tracer version 1.5 and the tree information was summarized using TreeAnnotator version 1.5.2 with a burn-in of 30,000 (20,000 trees summarized) (all software included with BEAST). However, intraspecific data evolve at genealogical scales and the inappropriate use of deep fossil calibration points that reflect phylogenetic time-scales can lead to misleading, usually overestimated, divergence times (Ho et al. 2008).

Additionally, Bayesian Skyline Plots (BSPs) were calculated for some lineages (those with > 20 samples) and continental groups using BEAST. The BSP estimates the posterior distribution of effective population size through time from a sample of gene sequences, given a specified evolution model, using a standard MCMC (Drummond et al. 2005). The analyses for each lineage were run for at least 60 million generations, sampling every 1000, using a piecewise linear BSP model (which assumes that effective population size is constant between coalescent

events but may change at different coalescent event times) with the appropriate evolutionary model (estimated with jModelTest), adjusting the operators and visualizing the results in Tracer. The 95% Highest Posterior Density (HPD) was included in the TMRCAs and BSPs estimations.

The nuclear marker data (Y-chromosome introns and BRCA1 gene) were not used to estimate historical demographic parameters of populations because of the low number of informative sites and lack of variation in comparison to the mitochondrial data.

**2.3 Results**

2.3.1 Sequence polymorphism

The cyt *b* was the most informative of the markers analysed in this study, while the concatenated Y-chromosome introns showed more variation than the autosomal nuclear gene, BRCA1, which was the least informative.

A partial sequence of 1110 bp for the cyt *b* gene was obtained. I found 319 *S. minutus* haplotypes in the 541 sequences analysed (plus 3 haplotypes of *S. volnuchini*), from which 286 have been described in previous studies and 33 are newly described here (Appendix 3). The full dataset presented a total of 412 mutations, 736 invariable sites and 374 polymorphic sites, from which 266 were parsimony informative and 108 were singleton mutations. There were 277 synonymous and 67 non-synonymous changes for the full sample (while excluding the outgroup). No internal stop codons were found and there were no insertion/deletions that could change the reading frame of the gene, enough evidence to presume that the sequences obtained were not pseudogenes.

The MK test was not significant between *S. minutus* and the outgroup (G = 1.0290, *p* = 0.3103); therefore, I assumed that the cyt *b* gene is selectively neutral

for *S. minutus*, and that the genetic structure observed here was the result of demographic and historical processes.

For the Y-chromosome introns, I obtained a concatenated sequence of 1143 bp. There were only 26 variants in the 127 sequences analysed (including the outgroup), but there were only 21 variants when not considering the sites with alignment gaps (Appendix 3). Six sequence variants are newly described in this study. There were 30 sites with alignment gaps, most present in a single intronic region in two samples from Austria (AUT1 and ATDo1611), 1078 invariant sites and 35 polymorphic sites, from which 23 were parsimony informative and 12 were singleton mutations.

The sequence length of the partial BRCA1 autosomal gene was 750 bp. In the 37 sequenced samples (Appendix 3), there were 27 sequences with no ambiguous nucleotide sites, nine sequences with one ambiguous nucleotide site and one sequence with two ambiguous nucleotide sites. The ambiguous nucleotide sites were interpreted as heterozygous individuals. Each homozygous individual contributes one haplotype, each heterozygous individual contributes two haplotypes, and the individual with two ambiguous nucleotide sites contributes four haplotypes because I cannot determine the phase. This gave a total of 14 different haplotypes, characterized by 11 polymorphic sites, only 6 of which were parsimony informative, and there were four synonymous and seven non-synonymous substitutions.

2.3.2 Phylogenetic analyses

The molecular markers used in this study had different degrees of sequence information and gave different phylogenetic resolutions of the tree topologies. The phylogenetic methods (NJ, MP, ML and BA) were consistent, because they produced similar trees for each molecular marker. *Sorex minutus* formed a separate

monophyletic clade from *S. volnuchini* with any molecular marker and phylogenetic method, as expected. When considering the main branches only, the phylogenetic methods produced comparable, but not identical, tree topologies between cyt *b* and Y-chromosome introns, but the BRCA1 gene data was poorly informative and did not resolve a similar topology. The lack of a strong phylogeographic structure with nuclear markers does not contradict the mitochondrial findings, but show that the selected Y-chromosome introns and nuclear gene are not adequate for these kinds of intraspecific studies because of the very low genetic variation that they contain.

With the cyt *b* data, the phylogenetic reconstructions showed similar tree topologies and phylogroups as previously found with a more restricted number of samples (Bilton et al. 1998, Mascheretti et al. 2003) and another mitochondrial marker (McDevitt et al. 2010). This means that the cyt *b* gene contains a strong evolutionary signature that supports several distinct phylogenetic lineages. For this marker, I defined the ingroup (composed of all the *S. minutus* samples) as the 'Eurasian' clade, within which I detected several other well supported branches that corresponded to distinct phylogenetic groups or 'phylogroups'. I defined these phylogroups by the geographical origin of the samples as follows (Fig. 2.1, Fig. 2.2, and see Appendix 3 for the allocation of samples into phylogroups): 1) 'North-Central European', the most widespread phylogroup; it included samples from Central France and Britain, extended across Central and Northern Europe to Lake Baikal, Siberia, but did not include samples from Southern Europe. 2) 'North-Central Italian', this lineage was mostly restricted to the Italian peninsula; it included samples from the Apennines and the Alps in Italy, but also from Switzerland, Slovenia, Southern and Eastern France near the border with Italy, Czech Republic and one distant sample from Germany. 3) 'Pyrenean', with a widespread geographic distribution; it included samples from the Pyrenees in Southern France and Andorra, and from the Pyrenean and Cantabrian Mountains in Northern Spain, but also

samples from Western France, Ireland, the Orkney islands, several islands on the western coast of Britain and a few samples from the British mainland. 4) 'South Italian', geographically restricted to La Sila, Calabria in Southern Italy (four samples) and also distinguished with the Y-chromosome introns and BRCA1 gene datasets. 5) 'Iberian', geographically restricted to the Iberian peninsula; it included samples from Rascafría, Central Spain (Sierra de Guadarrama) and Picos de Europa, Northern Spain. 6) 'Balkan', that included samples from the Balkan peninsula, belonging to the European part of Turkey, Serbia, Montenegro, Macedonia and Bosnia Herzegovina, and samples from further north in Slovakia, Czech Republic, Austria and Switzerland; a similar lineage was found with the nuclear markers. There was a basal polytomy including the clade containing the 'Pyrenean', 'South Italian' and 'North-Central Italian' phylogroups, and the 'Balkan' and 'Iberian' phylogroups, and low bootstrap support for the clade containing those lineages.

Besides the previously defined main phylogroups, I also created the 'continental Eurasian', 'continental North-Central European' and 'continental Pyrenean' phylogroups extracting the island samples from the Eurasian, North-Central European and Pyrenean phylogroups, respectively. Also, within the North-Central European phylogroup I distinguished a phylogroup from 'Britain' (Fig. 2.2) composed of samples from England, Scotland and Wales, and islands off the coast of mainland Britain (excluding the Orkney islands and other island samples that belonged to the Pyrenean phylogroup). Within the Pyrenean phylogroup I distinguished three island groups based on geographical isolation and proximity (Fig. 2.2): the 'Ireland' phylogroup (a monophyletic clade), the 'Orkney' group (a genetically heterogeneous group of samples composed of Mainland Orkney, South Ronaldsay and Westray), and a 'Celtic fringe' group (composed of samples from islands off the coast of Western Britain and a small number of samples from mainland Britain that did not cluster with the North-Central European phylogroup;

Searle et al. 2009). These continental and island samples were analysed as separate groups because there could be a noticeable effect of the island samples on the estimation of some phylogeographic results.



**Figure 2.1.** Sample localities of *Sorex minutus* from Europe divided into cytochrome *b* phylogroups: ● North-Central European (three Russian samples from Siberia are not shown in map), ○ Pyrenean, ☆ South Italian, □ North-Central Italian, ▽ Balkan and ▲ Iberian; ■ *S. volnuchini* (outgroup). Relief illustrated in grey tones.

**Figure 2.2.** Bayesian inference tree of cytochrome *b* data showing the phylogenetic relationships among *Sorex minutus* haplotypes. Phylogroups (North-Central European, Pyrenean, South Italian, North-Central Italian, Balkan and Iberian) and four other groups (Britain, Ireland, Orkney and Celtic fringe) are represented with different colours. Letters 'A – F' correspond to ancestral (central) haplotypes in the MJ network (Fig. 2.3). Values on branches correspond to posterior probability support. The scale bar represents 0.3 nucleotide substitutions per site.

The MJ network also resolved six main phylogroups separated by several mutational steps (Fig. 2.3A). The Iberian, Pyrenean, Balkan and Northern-Central European groups were separated by >10 mutational steps and *S. volnuchini* samples were distantly connected to the Eurasian clade through the North-Central European phylogroup (44 steps).

The North-Central European phylogroup had a star-like pattern with three central haplotypes from where all other haplotypes derived (Fig. 2.3B): the most internal one ('A') was a haplotype with one sample from Central Ukraine (UATi0266), connected by one mutational step to the second central haplotype ('B') formed by two samples from the Netherlands (NLBo0001and NLZV0044), which was connected by one mutational step to the third central haplotype ('C') formed by four samples from Britain (GBAn0042, GBDo0459, GBKk0100 and GBNo0002) plus three sequences from the Netherlands (NLDi0001, NLFr0017 and NLFr0040). There was an apparent geographical subdivision within this North-Central European sub-network, where samples from Siberia, Eastern and Northern Europe were derived or most closely connected to 'A', samples from Central Europe were derived or most closely connected to 'B', and all samples from Britain were derived or most closely connected to 'C', than to the other central haplotypes. However, this geographical subdivision was weak because there were several loops among the three central haplotypes in the inner part of the North-Central European network, and only the British and Dutch samples that connected to the central haplotype 'C' formed a well supported group in the phylogenetic analyses. Loops in the networks represent alternative, equally parsimonious explanations of the relationships among haplotypes. Moreover, there were some highly divergent sequences from Russia (Lake Baikal in Siberia), Poland and Ukraine with more than 8 nucleotide differences separating them from 'A'.
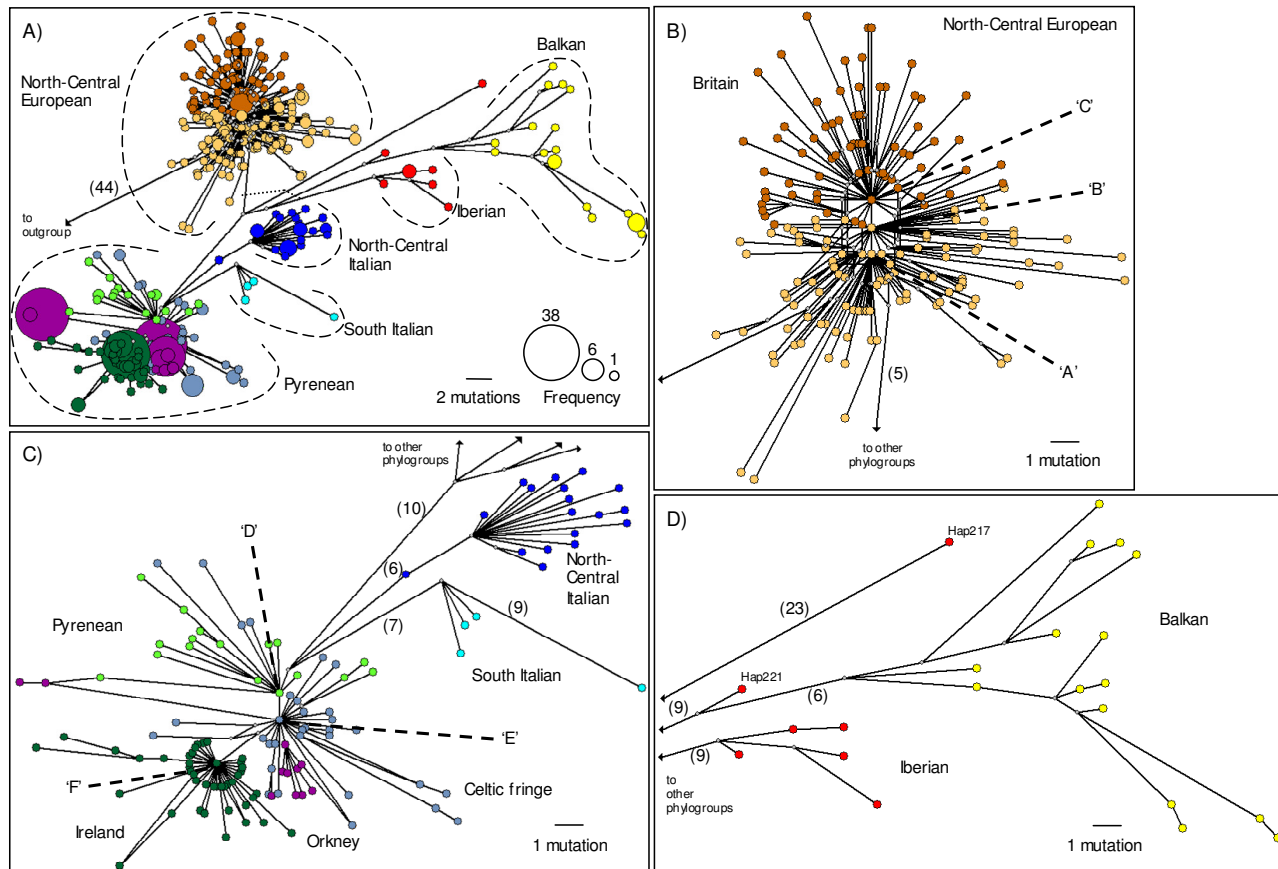
**Figure 2.3.** Median-Joining network of cytochrome *b* data showing the relationships among *Sorex minutus* haplotypes. A) Full network showing the main phylogroups (dashed lines). B) The North-Central European phylogroup (with British group). C) The Pyrenean (with Ireland, Orkney and Celtic fringe groups), North-Central Italian and South Italian phylogroups. D) The Balkan and Iberian phylogroups. Ancestral (central) haplotypes from where all other haplotypes derive from are marked with letters 'A – F' (see text for explanation). Numbers in parentheses represent mutation steps. Dotted line in centre of A) shows where the network was cut.

The Pyrenean phylogroup also had a star like pattern, but it was more complex than the North-Central European clade (Fig. 2.3C). The Pyrenean haplotype ESBa9709 was the most basal of this part of the network ('D'), and all other Pyrenean and some British samples (from the Celtic fringe) were derived from it by one or more mutational steps. Two island samples from the west coast of Britain (GBAr0002 and GBMu0858), sharing the same haplotype ('E'), were central to several other British island samples connected to them by one or more mutational steps, and all belonged to the Celtic fringe group. Notably, two samples from Navarra, Northern Spain plus 29 Irish samples formed a central haplotype ('F') for all Irish sequences, most of them connected by one mutational step only.

The North-Central Italian and South Italian phylogroups were most closely connected to the Pyrenean clade in the MJ network (Fig. 2.3C), in agreement with the phylogenetic trees. The North-Central Italian phylogroup showed no apparent geographical substructure within it and haplotypes were connected among each other by several mutational steps through one hypothetical vector. A sample from North-Western Italy (ITGe0003) was separated by a few steps from the rest of the Italian samples. More samples from Southern Italy are needed for a better comprehension of the genetic structure there, but the current data showed clear differentiation from the North-Central Italian samples and substantial variation among the specimens from La Sila Mountain, with one sample (ITSi0017) separated by 9 mutational steps from the rest.

The Balkan phylogroup (Fig. 2.3D) showed no star like pattern and the haplotypes from this region formed two highly differentiated groups: one comprised samples from the southern parts of the Balkans (European Turkey, Macedonia, Montenegro and Serbia), and the other one comprised samples from the southern and central parts of the Balkans (Macedonia, Montenegro and Bosnia-Herzegovina) plus Austrian, Slovak, Czech and Swiss samples. This phylogroup displayed the

highest number of mutational steps within a group, i.e., making it the most diverse, in agreement with the phylogenetic trees.

The Iberian sub-network was represented by six haplotypes only (Fig. 2.3D); further sampling in this area is needed to infer the genetic structure there, but these results indicate high genetic diversity and differentiation.

The main phylogroups were geographically distinct but there were areas of overlap with other phylogroups, termed contact zones. For example, samples from Czech Republic and Switzerland belonged to the North-Central Italian, the Balkan or the North-Central European phylogroups, so a contact zone among these phylogroups in Central Europe can be inferred; samples from Central France belonged to the Pyrenean or the North-Central European phylogroups, so a contact zone can also be inferred somewhere in Central France; samples from Austria and Slovakia belonged to the Balkan or the North-Central European phylogroups, showing that there is a contact zone north of the Balkan peninsula (Fig. 2.1, and see Appendix 3 for the allocation of samples into phylogroups).

Several samples had particularly interesting haplotypes. Two samples from Navarra, Northern Spain (ESNa0861 and ESNa1131) clustered with and had the same cyt *b* haplotype as several Irish samples. South Italian samples formed a monophyletic group that clustered with the Pyrenean phylogroup with any phylogenetic method used, but not with the geographically closer North-Central Italian phylogroup. The Iberian and Balkan phylogroups had the most diverse sequences, and the later had two subclades, as explained in the network results, with an apparent distinction of southern and northern samples from the Balkan peninsula.

There were three problematic sequences: one sample from Rascafria, Central Spain (ESRa0001) clustered separately from other Iberian samples and it was quite divergent from all other sequences, and another one from the same area

(ESRa3448) clustered within the Balkan phylogroup. One Central Italian sample from Cerreto Laghi (ITLC5646) always clustered within the Iberian phylogroup. Long distance migration events could explain these strange patterns, but this is highly unlikely for *S. minutus*. More likely is that these three samples were either mislabelled or contaminated during collection or during processing. ESRa0001 and ITLC5646 are haplotype sequences obtained from GenBank and I did not have the original samples to verify this. Degraded genomic DNA from ESRa3448 was obtained from brain tissue of a skull sample, but the chromatograms showed a good sequencing signal and I have no reason to believe it could have been contaminated with DNA from Balkan samples, so I suppose that it has been mislabelled. Another possible explanation is that these samples are representative of ancient haplotypes, still present in the Eurasian sample, relating to an older colonisation event. As a precaution, these samples were not included in the phylogroup sequence diversity estimations.

The phylogenetic reconstructions were less resolved with the Y-chromosome introns than with the cyt *b* data because of the low number of parsimony informative sites (Fig. 2.4A, and see Appendix 3 for the allocation of samples into phylogroups). All phylogenetic methods distinguished a 'North-Central Italian' group, represented by one haplotype only (with samples CHBr5421, ITMa0179, IT533, IT56, SWZ47 and SWZ54, but also including sample ITLC5646 which in the cyt *b* phylogeny clustered with the Iberian samples), a 'South Italian' group, represented by one haplotype (samples ITCa5342, ITSi0017, ITSi0021 and IT534) all from Calabria, Southern Italy, and a 'Balkan' group, represented by three haplotypes from the southern Balkan peninsula (samples MEMK0001, RSMK7008 and TUR7). Moreover, the Pyrenean and Iberian lineages found with the cyt *b* data were indistinguishable from each other with the Y-chromosome data, and shared the same sequence or clustered with North-Central European samples, forming the

geographically widespread lineage 'Pyrenean + Iberian + North-Central European'. Also, several North-Central Italian and Balkan samples, which in the cyt *b* data clustered separately, formed an unresolved group with the Y-chromosome intron data named 'North-Central European + North-Central Italian + Balkan'.

For the BRCA1 gene (Fig. 2.4B), the only group that showed bootstrap support was composed of Pyrenean, Orkney and Iberian samples. A South Italian clade was resolved with different phylogenetic methods, but it was poorly supported.
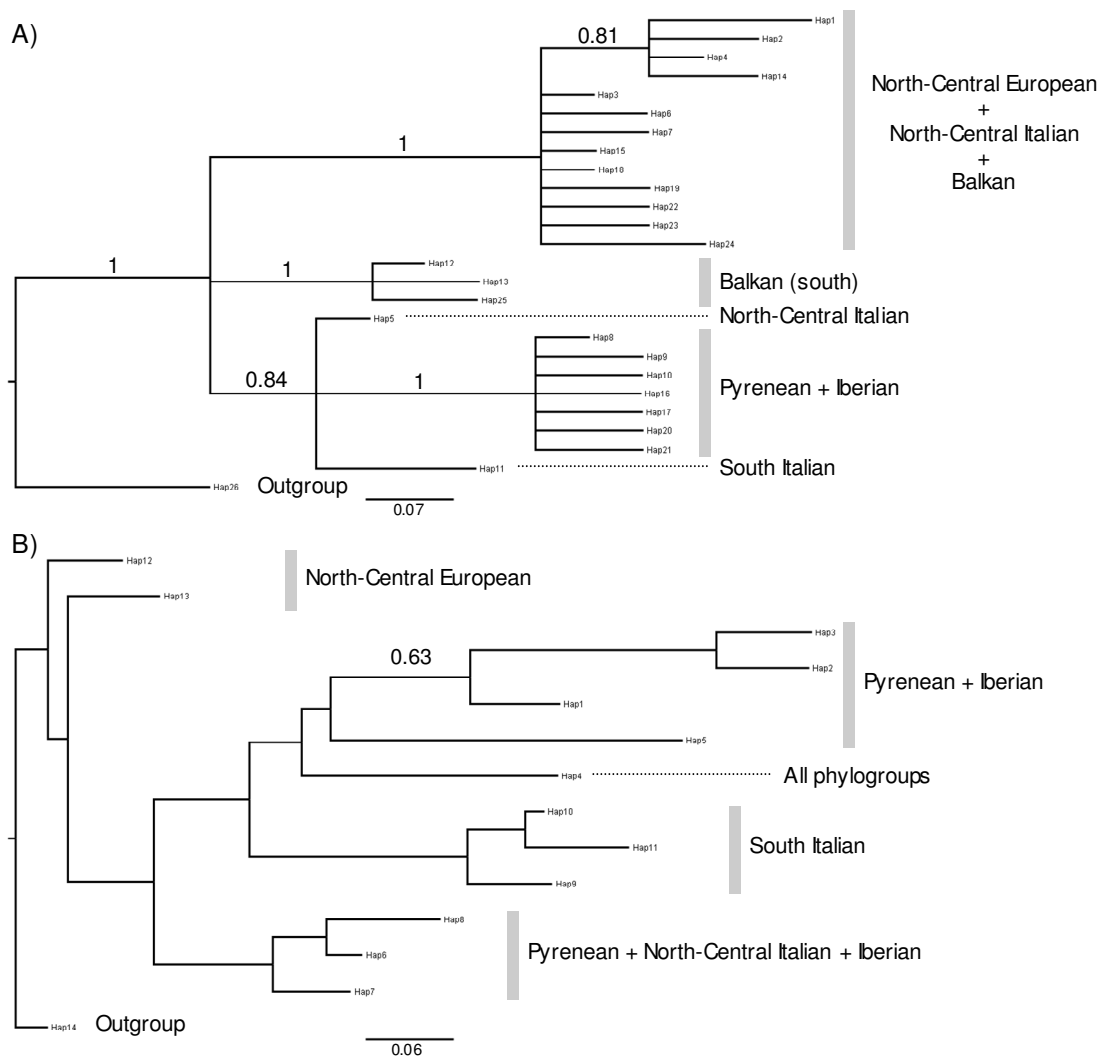


**Figure 2.4.** Bayesian inference tree of A) Y-chromosome introns and B) BRCA1 gene showing the phylogenetic relationships among *Sorex minutus* sequences. Values on branches represent posterior probability branch support.

2.3.3 Sequence diversity

Nucleotide diversity ($\pi$) and haplotype diversity (Hd) for the Eurasian sample and for the phylogroups found with the cyt *b* gene was high and similar to those found in previous studies (Table 2.2). Nucleotide diversity values for the Eurasian and continental Eurasian samples were an order of magnitude higher than for the phylogroups. The phylogroup with the highest nucleotide diversity was the Balkan ($\pi_{Balkan}$ = 0.0099), as expected for the highly divergent sequences there, while the smallest diversity values were found in the island groups, Ireland and the Orkney islands ($\pi_{Ireland}$ = 0.0020, $\pi_{Orkney}$ = 0.0027), as expected for the low number of haplotypes and genetic similarity among samples. There was little effect of island genetic diversity on the total diversity estimations, although the number of haplotypes was rather different, because island samples were usually represented by very similar or identical sequences (Table 2.2). For example, the Eurasian sample had a slightly smaller nucleotide diversity ($\pi_{Eurasian}$ = 0.0147) than the continental Eurasian sample (excluding samples from Britain, Ireland and other islands; $\pi_{cont-Eurasian}$ = 0.0171); likewise, the nucleotide diversities of the North-Central European ($\pi_{NCEuropean}$ = 0.0065) and continental North-Central European ($\pi_{cont-NCEuropean}$ = 0.0065), and those of the Pyrenean ($\pi_{Pyrenean}$ = 0.0049) and continental Pyrenean ($\pi_{cont-Pyrenean}$ = 0.0050) groups were identical. The Celtic fringe group had high diversity values compared to other groups of the Pyrenean phylogroup ($\pi_{Celtic-fringe}$ = 0.0057, $Hd_{Celtic-fringe}$ = 0.9710), but this was the result of pooling samples from different localities.

Sequence diversity for *S. minutus* Y-chromosome introns and BRCA1 gene was $\pi$ = 0.0046 and $\pi$ = 0.0015, respectively. Sequence diversity for the main phylogroups of the Y-chromosome introns and nuclear gene was not computed because of the low number of polymorphic sites.

**Table 2.2.** Cytochrome *b* sequence diversity of phylogroups, island and continental samples of *Sorex minutus*.

| Phylogroup | n | Haplotypes | Hd | π | K | S |
|---|---|---|---|---|---|---|
| Eurasian | 541 | 319 | 0.9847 | 0.0147 | 16.1190 | 362 |
|    Continental Eurasian | 177 | 148 | 0.9974 | 0.0171 | 18.6990 | 252 |
| North-Central European | 198 | 172 | 0.9977 | 0.0065 | 7.1800 | 253 |
|    Continental North-Central European | 102 | 90 | 0.9970 | 0.0065 | 7.1440 | 170 |
|    Britain | 80 | 69 | 0.9950 | 0.0054 | 5.9510 | 132 |
| Pyrenean | 283 | 102 | 0.9458 | 0.0049 | 5.4400 | 147 |
|    Continental Pyrenean | 15 | 13 | 0.9810 | 0.0050 | 5.5430 | 28 |
|    Ireland | 96 | 42 | 0.8840 | 0.0020 | 2.1730 | 53 |
|    Orkney | 119 | 11 | 0.7720 | 0.0027 | 3.0140 | 17 |
|    Celtic fringe | 50 | 33 | 0.9710 | 0.0057 | 6.2600 | 71 |
| North-Central Italian | 26 | 18 | 0.9600 | 0.0061 | 6.7720 | 51 |
| South Italian | 4 | 4 | 1.0000 | 0.0073 | 8.0000 | 16 |
| Balkan | 21 | 16 | 0.9570 | 0.0099 | 10.9050 | 53 |
| Iberian | 7 | 5 | 0.8570 | 0.0052 | 5.7140 | 15 |

n = Number of sequences.

Hd = Haplotype diversity.

π = Nucleotide diversity (Jukes and Cantor 1969).

K = Average number of nucleotide differences.

S = Number of polymorphic (segregating) sites.

2.3.4 Divergence and genetic structure of populations

In general, there were small pairwise sequence divergence (Da) values among phylogroups with the cyt *b* data, consistent with intraspecific studies (Table 2.3). The smallest divergence was between the Pyrenean and the South Italian phylogroups (Da = 0.7%), followed by the divergence between the Pyrenean and the North-Central Italian phylogroups (Da = 0.9%). Interestingly, the South Italian and North-Central Italian phylogroups presented a larger divergence (Da = 1.0%) despite the geographical proximity in the Italian peninsula, in accordance with the phylogenetic tree and MJ network. The largest divergences were found among the Balkan and the other phylogroups (Da values between 1.7% and 2.4%), followed by the divergences among the Iberian and the other phylogroups (Da values between

1.7% and 2.2%). About 5.2% sequence divergence, on average, was found between the ingroup and the outgroup (average Da = 0.0519) as expected for interspecific sequence divergence.

Pairwise genetic differentiation values among phylogroups based on $F_{ST}$ were large and significant (Table 2.4). Again, the smallest differentiation value was between the Pyrenean and the South Italian phylogroups ($F_{ST}$ = 0.5922). The greatest differentiation was between the Pyrenean and the Balkan phylogroups ($F_{ST}$ = 0.8120), and between the Pyrenean and the Iberian phylogroups ($F_{ST}$ = 0.8081). However, I found no correlation between genetic divergence and geographic distances ($R$ = 0.3871, $p$ = 0.1313), or between genetic differentiation and geographic distances among phylogroups ($R$ = 0.0618, $p$ = 0.4061); therefore, there was no evidence of isolation by distance at the level of phylogroups.

**Table 2.3.** Pairwise sequence divergence values Da (SD), below diagonal, and pairwise geographic distances (in Km), above diagonal, among cytochrome *b* phylogroups of *Sorex minutus*.

| | Pyrenean | South Italian | North-Central Italian | North-Central European | Iberian | Balkan | Outgroup |
|---|---|---|---|---|---|---|---|
| Pyrenean | - | 2385 | 1616 | 843 | 1554 | 2043 | 3103 |
| South Italian | 0.0069 (0.0018) | - | 773 | 1698 | 1764 | 621 | 1307 |
| North-Central Italian | 0.0090 (0.0008) | 0.0102 (0.0029) | - | 942 | 1350 | 553 | 1695 |
| North-Central European | 0.0142 (0.0004) | 0.0148 (0.0016) | 0.0153 (0.0007) | - | 1597 | 1250 | 2265 |
| Iberian | 0.0206 (0.0028) | 0.0213 (0.0080) | 0.0223 (0.0041) | 0.0168 (0.0014) | - | 1884 | 2987 |
| Balkan | 0.0225 (0.0018) | 0.0239 (0.0059) | 0.0235 (0.0027) | 0.0187 (0.0010) | 0.0174 (0.0040) | - | 1145 |
| Outgroup | 0.0524 (0.0127) | 0.0524 (0.0262) | 0.0508 (0.0150) | 0.0486 (0.0093) | 0.0529 (0.0238) | 0.0540 (0.0168) | - |

**Table 2.4.** Pairwise $F_{ST}$ values (below diagonal) among cytochrome *b* phylogroups and phylogroup specific $F_{ST}$ values (diagonal) of *Sorex minutus*.

| | Pyrenean | South Italian | North-Central Italian | North-Central European | Iberian | Balkan |
|---|---|---|---|---|---|---|
| Pyrenean | 0.7210 | | | | | |
| South Italian | 0.5922 | 0.7206 | | | | |
| North-Central Italian | 0.6446 | 0.6235 | 0.7202 | | | |
| North-Central European | 0.7191 | 0.6964 | 0.7047 | 0.7198 | | |
| Iberian | 0.8081 | 0.7721 | 0.7869 | 0.7213 | 0.7211 | |
| Balkan | 0.8120 | 0.7132 | 0.7517 | 0.7352 | 0.6559 | 0.7175 |

The overall pattern of intraspecific genetic differentiation was also strong and highly significant ($p < 0.0001$), about 72% of the variation was found among phylogroups (Table 2.5).

**Table 2.5.** Analysis of Molecular Variance (AMOVA) among cytochrome *b* phylogroups of *Sorex minutus* and phylogroup $F_{ST}$.

| Source of variation | DF | Sum of squares | Variance components | Percentage of variation |
|---|---|---|---|---|
| Among populations | 5 | 2595.872 | 8.2139 | 72.04 |
| Within populations | 532 | 1696.01 | 3.1880 | 27.96 |
| Total | 537 | 4291.881 | 11.4019 | 100 |
| Overall (Eurasian)  $F_{ST} = 0.7204$ ($p < 0.001$) | | | | |

DF = Degrees of Freedom

Using the molecular clock, all intraspecific divergence times fell approximately between 0.2 and 0.6 MYA, within the Middle Pleistocene (Table 2.6). The divergence times among the outgroup and the phylogroups fell between 1.2 and 1.3 MYA, within the Early Pleistocene (Table 2.6), but this divergence time could be underestimated because of substitutional saturation or low sample size for the outgroup. The Balkan and Iberian phylogroups diverged approximately 0.4 MYA, and both had the largest values of divergence time compared to the other phylogroups ranging between 0.4 – 0.6 MYA. As expected from the lower divergence values, the South Italian and Pyrenean phylogroups diverged more recently, about 0.173 MYA, while the North-Central Italian and South Italian split about 0.255 MYA.

**Table 2.6.** Time of divergence (±SD) among cytochrome *b* phylogroups of *Sorex minutus* (in MYA).

| | Pyrenean | South Italian | North-Central Italian | North-Central European | Iberian | Balkan | Outgroup |
|---|---|---|---|---|---|---|---|
| Pyrenean | - | | | | | | |
| South Italian | 0.173 (0.127-0.218) | - | | | | | |
| North-Central Italian | 0.225 (0.205-0.245) | 0.255 (0.182-0.328) | - | | | | |
| North-Central European | 0.355 (0.345-0.365) | 0.370 (0.331-0.409) | 0.383 (0.366-0.399) | - | | | |
| Iberian | 0.515 (0.446-0.584) | 0.533 (0.332-0.733) | 0.558 (0.454-0.661) | 0.420 (0.385-0.455) | - | | |
| Balkan | 0.563 (0.518-0.607) | 0.598 (0.450-0.746) | 0.588 (0.519-0.656) | 0.468 (0.442-0.493) | 0.435 (0.335-0.536) | - | |
| Outgroup | 1.310 (0.993-1.627) | 1.310 (0.656-1.964) | 1.270 (0.896-1.644) | 1.215 (0.982-1.448) | 1.323 (0.729-1.916) | 1.350 (0.931-1.769) | - |

The genetic relationships among phylogroups are better visualised in the NJ tree using the pairwise distance values (Fig. 2.5): the Iberian and Balkan formed a divergent clade, separated from the North-Central European, North-Central Italian, South Italian and Pyrenean lineages; the Pyrenean, South Italian and North-Central Italian phylogroups clustered together, separated from North-Central European samples. These results showed a similar topology to the one found with the BA of cyt *b* haplotype sequences, the only difference was in the base of the NJ tree where the North-Central European lineage grouped with the Pyrenean and Italian lineages, and was not basal as in the Bayesian tree.
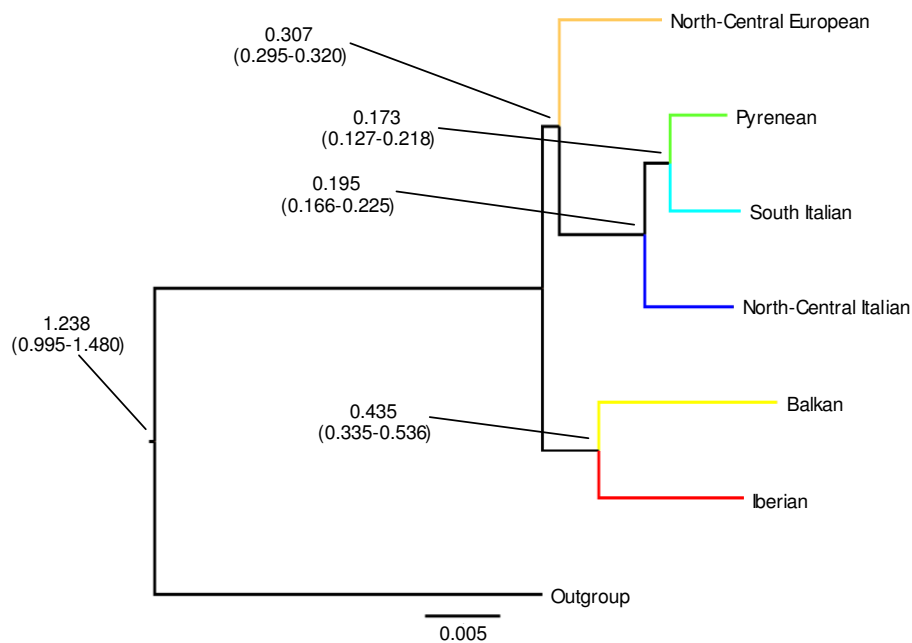


**Figure 2.5.** Neighbor-Joining tree showing the pairwise distances among cytochrome *b* phylogroups of *Sorex minutus* and estimated divergence times (in MYA) assuming a molecular clock.

I also calculated the genetic divergence and divergence times between the main clades observed on the NJ tree (Fig. 2.5). The genetic divergence between the 'Pyrenean + South Italian' and 'North-Central Italian' lineages was 0.8% (Da = 0.0078, divergence time = 0.195 MYA), between the 'Pyrenean + Italian peninsula'

and 'North-Central European' was 1.2% (Da = 0.0123, divergence time = 0.307 MYA), and between the 'Pyrenean + Italian peninsula + North-Central European' and 'Balkan + Iberian' was 1.4% (Da = 0.0136, divergence time = 0.340 MYA).

The TMRCA for each lineage are shown in Fig. 2.6. The values differed from the time of divergence estimates assuming a molecular clock, as expected from the different methodologies. The root of the Bayesian tree had a mean value of 8.039 MYA (95% HPD: 4.734 – 12.624 MYA), which marked the evolution of *S. minutus* from its ancestral population and split from its sister species *S. volnuchini*. The deepest node for *S. minutus* had an estimated TMRCA of 5.139 MYA (95% HPD: 4.532 – 5.312), in the Early Pliocene, congruent with the fossil calibration.

For each phylogroup the TMRCA ranged between 0.7 – 1.7 MYA. The TMRCAs involving multiple phylogroups tended to be much higher, ranging from 1.4 MYA (Pyrenean and South Italian phylogroups) to 5.1 MYA (North-Central European phylogroup and the clade formed by all other phylogroups). Because of the large 95% HPD in the Bayesian analysis, I was not able to estimate with confidence the TMRCAs for the island groups.
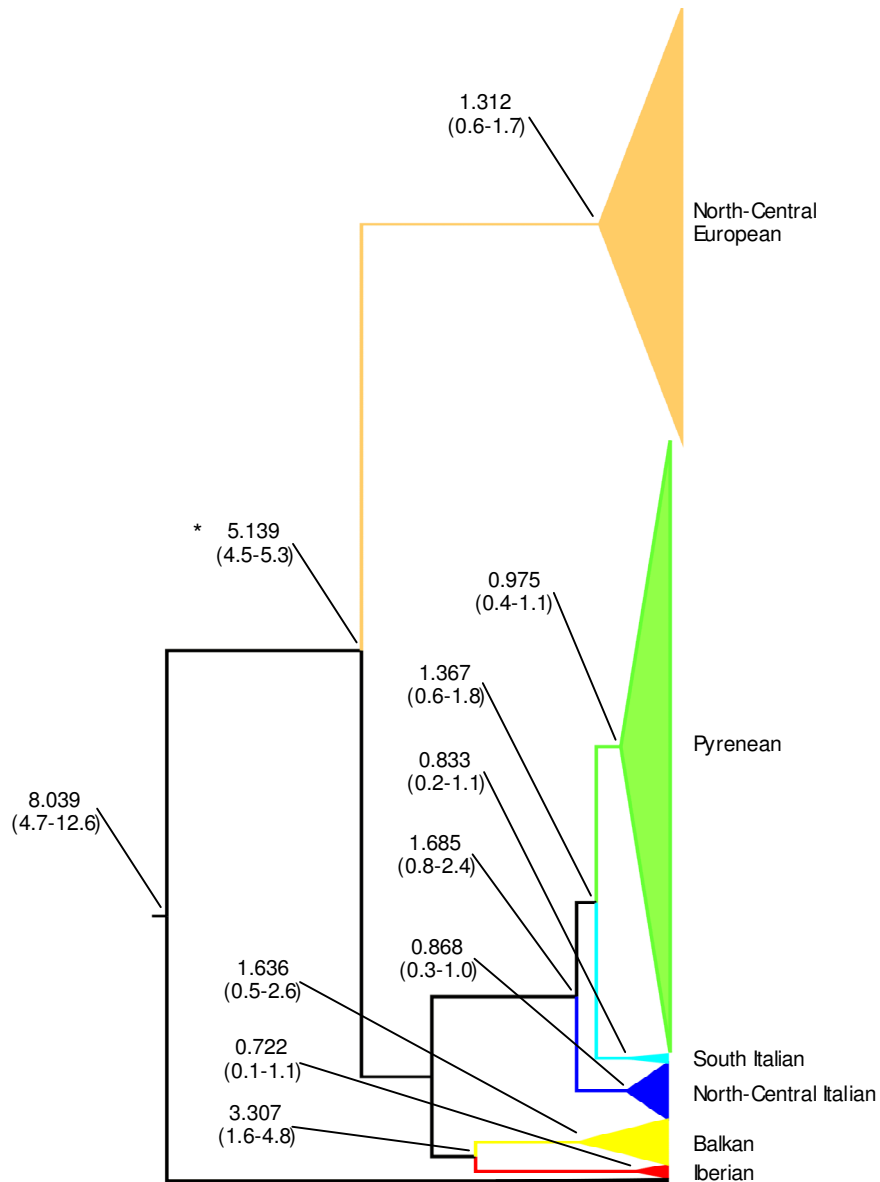
**Figure 2.6.** Bayesian inference tree of cytochrome *b* data showing the genetic relationships among phylogroups of *Sorex minutus*. Values on nodes correspond to Times to the Most Recent Common Ancestor with 95% Highest Probability Density confidence intervals (in MYA). The ingroup (*) was calibrated with the oldest known fossil record of *S. minutus* (5 MYA, SD = 0.2).

2.3.5 Historical demography

The mismatch distributions for the Eurasian and continental Eurasian samples were multimodal, but related to pairwise differences within and between phylogroups and not to an historically constant population size (Fig. 2.7). The

Pyrenean, North-Central European, North-Central Italian, continental Pyrenean, continental North-Central European, Ireland and Britain groups all had distinctly unimodal mismatch distributions with an almost perfect fit between observed and expected pairwise differences of a sudden population expansion model (Fig. 2.7).

Population expansion tests performed on the main phylogroups, continental and island groups are presented in Table 2.7. Because of low sample size, the population expansion tests were not performed on the South Italian and outgroup lineages (n < 5). The Eurasian and continental Eurasian samples had significant $R_2$ values and showed significant departures from neutrality with highly negative Fu's Fs and Tajima's D values.

There was consistency among the significant population expansion tests and unimodal mismatch distributions: the Pyrenean, North-Central European, North-Central Italian, continental Pyrenean, continental North-Central European, Ireland and Britain groups also had significant *p* values for $R_2$, Fu's Fs and Tajima's D statistics. The Balkan and Iberian phylogroups had non-significant *p* values for $R_2$, Fu's Fs and Tajima's D population expansion statistics, consistent with the multimodal mismatch distributions and the lack of fit between observed and expected pairwise difference distributions for a sudden population expansion model (Fig. 2.7). With the SSD test, all phylogroups showed non-significant differences from the modelled distribution under a sudden population expansion, likely reflecting the low statistical power of this test.

Interestingly, samples from the Orkney islands displayed a skewed mismatch distribution (Fig. 2.7), with a high frequency of zero pairwise differences (23%), because of the low number of nucleotide polymorphisms regardless of the large number of sequences obtained from there. This related to no evidence for population expansion based on the *p* values for Fu's Fs, Tajima's D and $R_2$ statistics (Table 2.7).
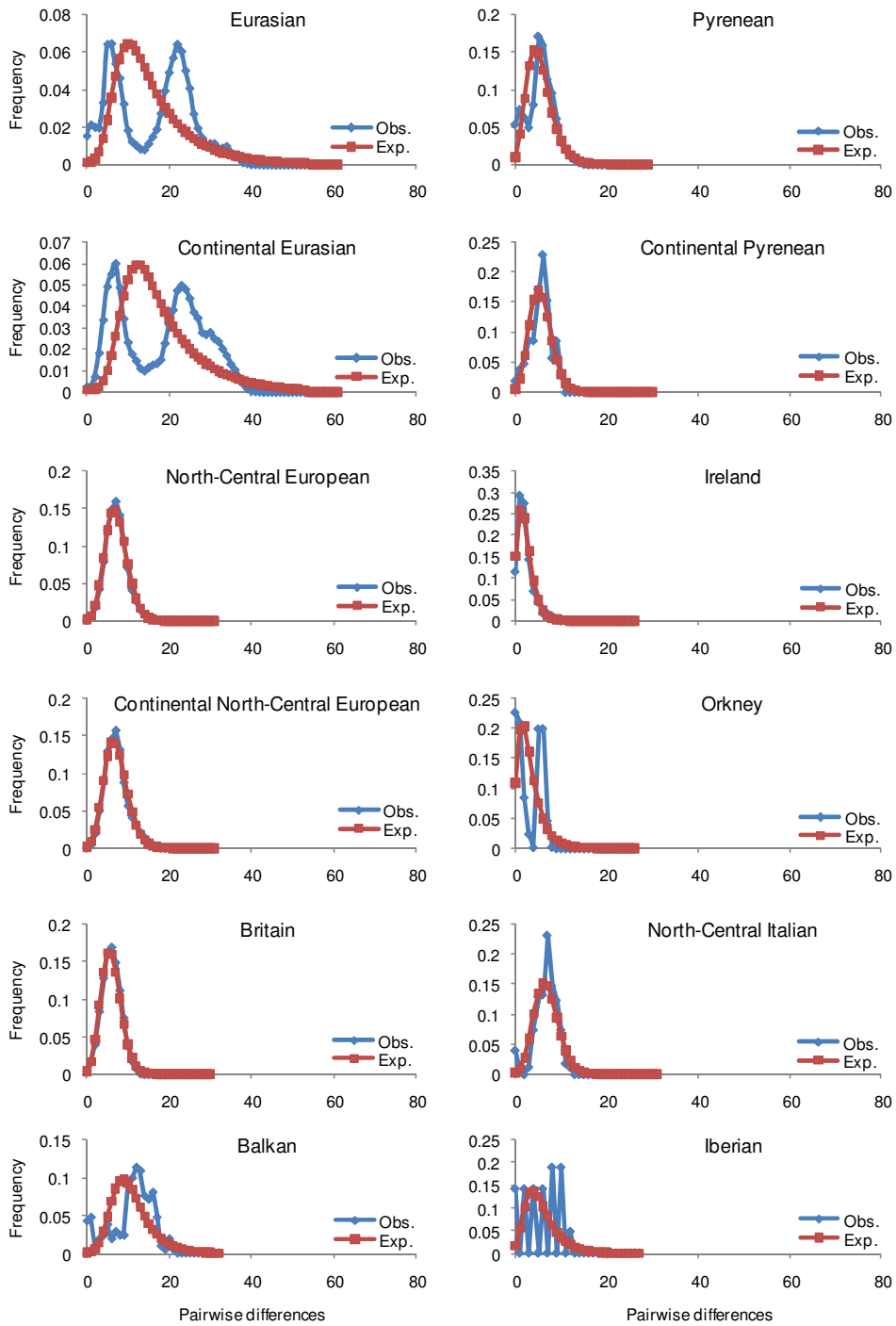
**Figure 2.7.** Mismatch distribution (frequency of pairwise nucleotide differences) of cytochrome *b* phylogroups of *Sorex minutus*.

**Table 2.7.** Population expansion analyses for the whole cytochrome *b* dataset (Eurasian) and phylogroups of *Sorex minutus* (*p* values are given between parentheses).

| Phylogroup | n | $R_2$ | Fu's Fs | Tajima's D | SSD |
|---|---|---|---|---|---|
| Eurasian | 538 | 0.0218 (< 0.01) | -457.6680 (< 0.01) | -2.1505 (< 0.01) | 0.0127 (0.3309) |
| Continental Eurasian | 174 | 0.0354 (< 0.01) | -159.6300 (< 0.01) | -1.9354 (< 0.01) | 0.0067 (0.5640) |
| North-Central European | 198 | 0.0123 (< 0.01) | -34.1384 (< 0.01) | -2.6878 (< 0.01) | 0.0004 (0.3119) |
| Continental North-Central European | 101 | 0.0172 (< 0.01) | -132.5150 (< 0.01) | -2.6491 (< 0.01) | 0.0005 (0.4694) |
| Britain | 92 | 0.0169 (< 0.01) | -97.6109 (< 0.01) | -2.6501 (< 0.01) | 0.0008 (0.3758) |
| Pyrenean | 283 | 0.0175 (< 0.01) | -114.6990 (< 0.01) | -2.3760 (< 0.01) | 0.0056 (0.2807) |
| Continental Pyrenean | 15 | 0.0793 (< 0.01) | -6.0342 (< 0.01) | -1.4937 (0.0596) | 0.0098 (0.4469) |
| Ireland | 96 | 0.0184 (< 0.01) | -52.7758 (< 0.01) | -2.5291 (< 0.01) | 0.0030 (0.1694) |
| Orkney | 119 | 0.0880 (0.5090) | 0.6044 (0.6515) | -0.1417 (0.5090) | 0.054 8 (0.1518) |
| North-Central Italian | 26 | 0.0521 (< 0.01) | -5.8766 (0.0159) | -1.8871 (0.0149) | 0.0119 (0.0744) |
| South Italian | 4 | NA | NA | NA | NA |
| Balkan | 21 | 0.0880 (0.0777) | -2.9944 (0.1050) | -1.0372 (0.1467) | 0.0067 (0.6385) |
| Iberian | 7 | 0.1396 (0.0847) | 0.5529 (0.5539) | -0.3685 (0.3779) | 0.0767 (0.3097) |
| Outgroup | 3 | NA | NA | NA | NA |

n = Number of samples

NA = Non-applicable (low sample size).

SSD = Sum of Square Deviations test.

The population expansion times calculated for those groups with a unimodal mismatch distribution and significant *p* values for $R_2$ and Fu's Fs statistics were older than the LGM (Table 2.8). However, for the Irish group the time of population expansion dated to 14 or 22 KYA (17 – 27 KYA), depending on the method used, close to the LGM. The North-Central European, Pyrenean and North-Central Italian phylogroups (and the continental and island counterparts) had expansion times between 41 and 81 KYA, all dated in the Late Pleistocene. Usually, the τ values (with confidence intervals) and expansion times obtained with Schneider and Excoffier's (1999) approach were similar to those obtained with Rogers' (1995) method; however, considerable differences in the estimation of τ and population

expansion times were found on the continental Eurasian and Pyrenean phylogroups (Table 2.8).

**Table 2.8.** Time of expansion parameters for the whole cytochrome *b* dataset (Eurasian) and phylogroups of *Sorex minutus*.

| Phylogroup | $\tau^1$ | Time | $\tau^2$ | Time |
|---|---|---|---|---|
| Eurasian | 7.4810 | 0.084 | 3.4 (1.107-23.873) | 0.038 (0.012-0.269) |
| Continental Eurasian | 9.3210 | 0.105 | 25.0 (14.336-32.381) | 0.282 (0.161-0.365) |
| North-Central European | 6.9730 | 0.079 | 7.2 (6.555-7.607) | 0.081 (0.074-0.086) |
| Continental North-Central European | 6.0930 | 0.069 | 6.6 (6.008-7.885) | 0.074 (0.068-0.089) |
| Britain | 5.9510 | 0.067 | 6.3 (5.379-6.885) | 0.071 (0.060-1.078) |
| Pyrenean | 3.6660 | 0.041 | 6.5 (3.723-8.676) | 0.073 (0.042-0.098) |
| Continental Pyrenean | 5.5430 | 0.062 | 6.1 (3.678-7.936) | 0.069 (0.041-0.089) |
| Ireland | 1.2420 | 0.014 | 2.0 (1.508-2.393) | 0.023 (0.017-0.027) |
| North-Central Italian | 6.7720 | 0.076 | 7.2 (5.715-8.494-) | 0.081 (0.064-0.096) |

[1] Based on Rogers (1995).

[2] Based on Schneider and Excoffier (1999).

95% confidence intervals are given between parentheses.

The traces of the BSPs converged, but had large variations of the posterior probability values and large 95% HPD, diminishing the confidence of this analysis. Figure 2.8 shows the results of the BSP of several lineages. Only the Pyrenean and Ireland groups had a notable increase of effective population size through time, while the other groups had more stable effective population sizes. The North-Central European had a relatively recent increase of population size, but the 95% HPD also increased. The group from Britain showed a stable population size, contrary to what was expected from the population expansion tests. However, the Balkan phylogroup had a stable, flat, effective population size through time, consistent with the non-significant results to the population expansion tests.
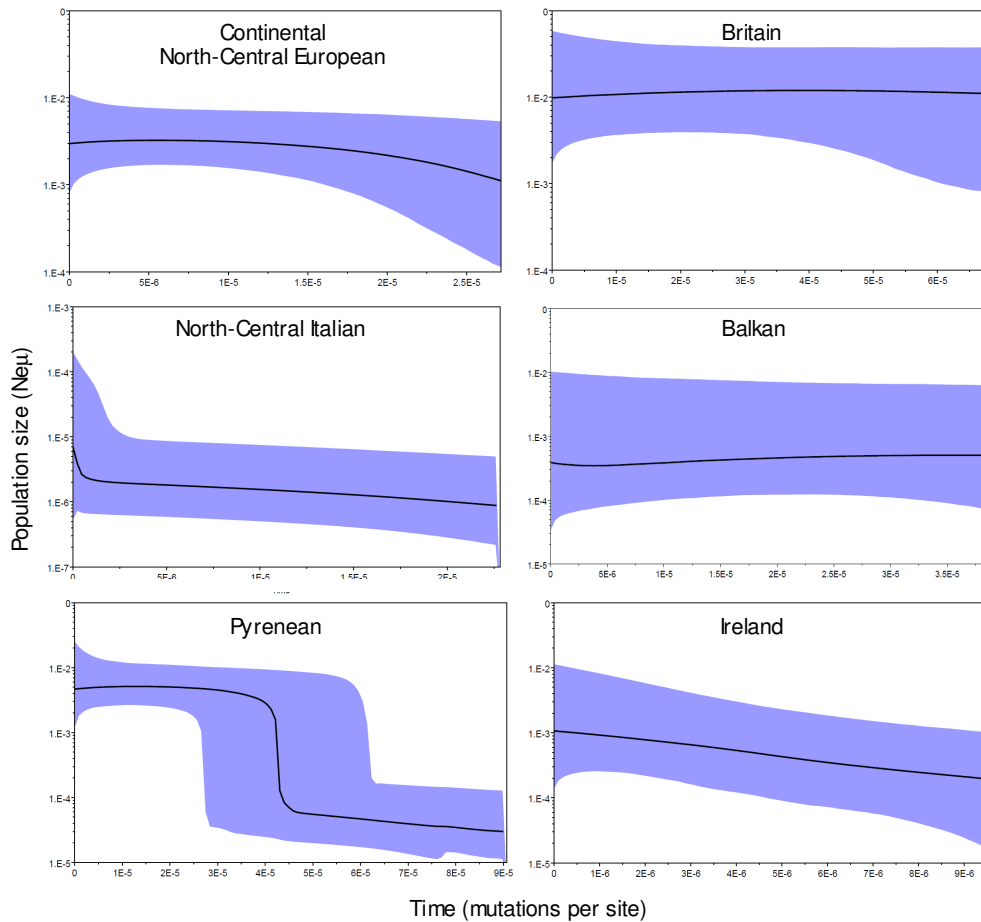
**Figure 2.8.** Bayesian Skyline Plots of cytochrome *b* phylogroups of *Sorex minutus* showing the historical demographic trends. The solid lines are median estimates and the shaded areas represent 95% Highest Probability Densities (confidence intervals).

## 2.4 Discussion and conclusions

The phylogeographic study that I present here notably expands previous findings on *S. minutus*, giving a more precise population genetic structure and demographic history, and contributing to the understanding of the patterns and processes during the Quaternary ice ages that shaped the European biota.

2.4.1 Phylogenetic findings and genetic population structure

The genetic lineages detected with cyt *b* data were similar to those previously found, either with fewer samples (Bilton et al. 1998, Mascheretti et al.

2003) or with different molecular markers (McDevitt et al. 2010). The Iberian, Balkan, North-Central Italian and South Italian phylogroups were geographically restricted to the Southern European peninsulas, consistent with the traditional Mediterranean glacial refugia; however, the North-Central European and Pyrenean phylogroups were distributed further north and represent evidence of northern refugia. The high genetic and haplotype diversities in southern and northern refugia reflect long term isolation and differentiation during the Pleistocene and Holocene, and the significant genetic structure illustrate a complicated history of European colonisation, not a simple case of isolation by distance and colonisation of Northern and Central Europe from expanding populations from the south.

It is interesting that the contact zones among phylogroups were detected at the northern extremes of the southern peninsulas. During the LGM, glaciers covered most of the Alpine and Pyrenean mountain ranges; therefore, these areas were deprived of fauna and flora, but when climate ameliorated and suitable habitat became available, pygmy shrew populations expanded and colonised previously glaciated areas only to find related populations from other lineages arriving from the other side of the mountain ranges, forming the contact zones among lineages that we observe today.

The poor phylogenetic resolution obtained with Y-chromosome introns and BRCA1 gene was not unexpected. Gündüz et al. (2007) found low divergence values among different species of ground squirrels with three Y-chromosome introns, ranging between 0.2% and 0.7% (compared to 5% – 10% with cyt *b*), and only eight single nucleotide polymorphisms and two insertion/deletions in 13 males analysed. Using the BRCA1 gene, Dubey et al. (2006) inferred the genetic relationships within the *Crocidura suaveolens* group and found several clades with low bootstrap support and basal polytomies resulting in partially unresolved phylogenies, but comparable to cyt *b* lineages. Moreover, the grouping of Pyrenean

and Iberian samples within a single clade was also found in another study using four concatenated Y-chromosome introns (McDevitt et al. 2010), and this could be related to the lower levels of nucleotide diversity in mammalian Y chromosomes caused by lower mutation rates compared to mtDNA (Hellborg and Ellegren 2004).

2.4.2 Insights on the colonisation routes into Europe

The origin of *S. minutus* from its proposed (extinct) ancestor, *S. minutoides*, has been estimated at the end of the Late Miocene in China (5.3 – 11.6 MYA; Storch et al. 1998, Rzebik-Kowalska 2005, 2008). The estimated time of evolution of *S. minutus* based on the Bayesian analysis with BEAST was consistent with this (estimated root at 8 MYA, ranging between 4.7 and 12.6 MYA). The oldest fossils of *S. minutus* were found in Podlesice and Mała Cave, Poland dated to the Early Pliocene between 4 and 5.3 MYA (Rzebik-Kowalska 1998), and an early widespread colonisation of Europe might have been possible because *S. minutus* was one of the first species of the genus *Sorex* in the continent (based on fossil records; Rzebik-Kowalska 1998, 2008). However, the very early separation of the different phylogroups from the Bayesian analysis with BEAST seems unrealistic compared to the molecular clock dating method I have used. The use of deep calibration points based on fossil records leads to slower observed mutational rates and overestimated evolutionary scenarios (Ho et al. 2008); therefore, these results should be taken with caution. Moreover, the use of a 5 million year calibration point for the earliest occurrence of *S. minutus* may be misleading, because that fossil found in Poland may have been an ancestor of both *S. minutus* and *S. volnuchini*, which are very similar to each other in morphology (see Chapter 4). Morphological analysis of the Polish fossils may be informative on this point.

The highly divergent sequences from Russia (Siberia), Poland and Ukraine, and the connection of *S. volnuchini* to the North-Central European phylogroup in the

MJ network argue in favour of a colonisation route of current *S. minutus* into Europe from the east or the south-east. A similar colonisation scenario before the LGM from Eastern to Western Europe was proposed for the common vole *Microtus arvalis*, which has a south-westward genetic gradient that could be a trace of the original colonisation of Europe (Heckel et al. 2005). A colonisation route originating from the east could also be hypothesised for other species that have the oldest divergence times with Ural phylogroups but have younger European phylogroups. For example, molecular dating suggests that the bank vole *Myodes* (*Clethrionomys*) *glareolus* diverged from a Ural phylogroup in the mid-Pliocene (2.98 MYA), but other European phylogroups diverged much more recently (Deffontaine et al. 2005); the North European and Central Asian groups of the root vole *M. oeconomus* diverged three to five glacial periods ago, but the divergence estimates for European groups were more recent (Brunhoff et al. 2003).

2.4.3 Colonisation and historical demography of southern peninsulas

The oldest fossils of *S. minutus* in the Balkans are dated to the Early Pliocene in Romania and Hungary between 4 and 4.5 MYA (Rzebik-Kowalska 1998), where the Balkan lineage is probably distributed, which could indicate the time of arrival of *S. minutus* into this southern peninsula. However, the estimated pairwise divergence times between the Balkan lineage and other phylogroups (within the last 0.6 MYA, Table 2.6) placed the divergence event much later.

The lack of a population expansion signature, high nucleotide and haplotype diversities, highly divergent sequences and apparent structure within the Balkan phylogroup suggest stable population sizes over time and genetic differentiation within the Balkan peninsula. The Balkans are a European hotspot for biodiversity given their environmental stability, topographic and climatic diversity and occasional connectedness with Asia Minor (Kryštufek and Reed 2004), and it could be

expected that some of these factors shaped the genetic diversity of the Balkan lineage there.

In the Iberian peninsula, the earliest fossil records of *S. minutus* are dated to the Early Pleistocene (1.2 – 1.4 MYA) (Cuenca-Bescós et al. 2010, Agustí et al. in press), again reflecting a long presence of *S. minutus* or a similar ancestral form in various part of Europe. Although Iberia is considered a refugial area during the Quaternary ice ages, and stable climate is suggested, it is a topographically diverse region with east-west mountain ranges and other high ground (over 1000 m.a.s.l.), large rivers (which could act as barriers to dispersal), and distinct seasonal precipitation and vegetation types (O'Regan 2008), which could have played an important role in the genetic differentiation of populations. The lack of significant population expansion values for the Iberian lineage may relate to historical stable population sizes; however, the sample size was low and this result should be taken with caution.

It is striking that the Pyrenean and South Italian lineages grouped together in the cyt *b* phylogenies and presented the smallest divergence times despite being geographically separated. This indicates a common history and it can be hypothesised that their common ancestor was more widespread throughout the Italian peninsula, probably displaced later by the North-Central Italian lineage in the Apennines and Western Alps. A similar scenario has been proposed for the water shrew *Neomys fodiens* (Castiglia et al. 2007), Alpine salamander *Salamandra salamandra* (Steinfartz et al. 2000), black pine *Pinus nigra* (Afzal-Rafii and Dodd 2007) and green lizard *Lacerta bilineata bilineata* (Böhme et al. 2007), which showed closely related South Italian and Pyrenean lineages but separated geographically by a North-Central Italian lineage. The Pyrenean phylogroup has significant signatures for population expansion, which suggest that this lineage, north of the traditional southern refugia, expanded during favourable climate

conditions. The expansion signature was not an effect of the island samples that belong to the Pyrenean lineage, because continental samples analysed separately also demonstrate a significant signature of population expansion.

The oldest fossils of the genus *Sorex* from Southern Italy, found in Basilicata region, are dated to the Early-Middle Pleistocene (ca. 0.5 – 1.2 MYA) and belong to an unrelated species, *S.* gr. *runtonensis-subaraneus* (Masini et al. 2005). Northern Italy also has shrew fossils, including a form classified as *S. minutus*, dated to the Early-Middle Pleistocene ca. 0.5 – 1.2 MYA, from Monte La Mesa near Verona (Masini and Sala 2007). The divergence times between the Pyrenean and South Italian lineages (about 0.173 MYA according to the molecular clock used, i.e., at the Middle-Late Pleistocene) suggest a much later arrival of the current population of *S. minutus* into Southern Italy. This lineage in the Calabria region is genetically and morphologically distinct from Central and North Italian populations, and this could have been a result of isolation and differentiation in an island setting, as hypothesised for other species and lineages in Southern Italy, reflecting island evolution (Vega et al. 2010a). Calabria consists of isolated mountain massifs separated by lowland areas in the southernmost part of the Italian peninsula, which from the Pliocene to the end of the Middle Pleistocene, at times of high sea level, were islands in a chain (Malatesta 1985, Caloi et al. 1989, Bonardi et al. 2001, Bonfiglio et al. 2002).

The finding of two genetic lineages of *S. minutus* in the Italian peninsula gives the impression that the refugial area may have been subdivided into multiple refugia at the LGM, one in the extreme south of Italy and at least one other refugium further north, concordant with the 'refugia within refugia' concept widely recognized for the Iberian peninsula (Gómez and Lunt 2007) and similar to microrefugia in the Balkans (Kryštufek et al. 2007). For the Italian peninsula, a similar 'refugia within refugia' pattern was found in a number of species (Vega et al. 2010a, and

references therein). The widespread North-Central Italian lineage may be presumed to derive from a glacial refugium located somewhere within the vicinity of the Apennine mountain chain. The significant population expansion signature for the North-Central Italian lineage demonstrates that it went through colonisation and expansion phases. However, further sampling and genetic analyses are needed to demonstrate population expansions from two separate refugial areas, and from the current data, it is possible that the North-Central and South Italian phylogroups occupied a single, continuous area at the LGM with the integrity of the groups retained by a hybrid zone (Barton and Hewitt 1985, Vega et al. 2010a).

2.4.4 Northern refugia of the North-Central European phylogroup

The high haplotype and nucleotide diversities and significant population expansion signatures of the North-Central European lineage are congruent with the hypothesis of northern glacial refugia (Bilton et al. 1998, Vega et al. 2010b). The BSP showed that the effective population size of this lineage remained either constant through time or had a steady increase of population size. If Mediterranean peninsulas were the source of northern populations, then the North-Central European phylogroup would not exist, there would be low genetic and haplotype diversities, concordant with a 'southern richness and northern purity' scenario (Hewitt 1996, 2000), and southern haplotypes would be found across Central and Northern Europe. Moreover, the star-like pattern in the phylogenetic network, the three ancestral haplotypes from Central and Eastern Europe identified by their central position from where peripheral, more recent, haplotypes derived, argue in favour of persistence and population expansion in that region (Avise 2000).

The hypothesis of northern refugia is further supported by palaeontological and palynological evidence for other temperate and boreal species (Willis et al. 2000, Willis and van Andel 2004, Magri et al. 2006, Sommer and Nadachowski

2006). Northern refugia have been hypothesised in phylogeographic analyses for a number of small mammal species, including the field vole *M. agrestis* (Jaarola and Searle 2002), bank vole *M. glareolus* (Deffontaine et al. 2005, Kotlík et al. 2006), root vole *M. oeconomus* (Brunhoff et al. 2003), common vole *M. arvalis* (Heckel et al. 2005) and common shrew *S. araneus* (Bilton et al. 1998, Yannic et al. 2008). Most of the phylogeographic studies point to the Carpathians as a likely northern refugial area for the North-Central European lineage, 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). However, for *M. glareolus*, *M. oeconomus*, *M. agrestis* and *M. arvalis*, and recently for *S. minutus*, predictions of their potential LGM distribution based on SDM are also consistent with northern refugia and indicate that the suitable climatic conditions could have been widespread across Central and Eastern Europe (Fløjgaard et al. 2009, Vega et al. 2010b).

Sorex minutus is considered a temperate species, but it is also latitudinally distributed above 60° North (i.e., near the Arctic Circle) and altitudinally above 2000 m.a.s.l. in regions with permafrost and harsh winters (Mitchell-Jones et al. 1999, Hutterer et al. 2008). Species that reach northern non-arctic distributions, 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), as 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. Moreover, *S. minutus* is currently restricted to high altitude areas in Southern Europe, because the dry, hot Mediterranean conditions in the lowlands do not suit it

(Taylor 1998). Therefore, its distribution in southern refugia may actually be more restricted during interglacials than glacials, in line with ideas of 'interglacial refugia' (Hilbert et al. 2007).

2.4.5 Colonisation history of Northern Europe and Britain

Populations of *S. minutus* belonging to the North-Central European lineage apparently colonised Northern Europe from refugia in Eastern-Central Europe. For example, Scandinavian and the Baltic regions were most likely colonised from Eastern Europe; thus, the phylogenetic network shows that sequences from Norway, Finland and Lithuania group closely with the Ukrainian ancestral haplotype. Central European populations likely originated from populations that remained in northern glacial refugia there. The stable population size over time and high genetic diversity indicate that the North-Central European lineage did not suffer from genetic bottlenecks during the glaciations.

From where did the British pygmy shrew populations arrive? The genetic similarity of samples from the Netherlands and Britain suggests that British *S. minutus* belonging to the North-Central European phylogroup originated from populations in the vicinity of the Netherlands (Vega et al. 2010b), reaching Britain over the Doggerland landbridge with continental Europe (Gaffney et al. 2007). Yet, there is a second lineage in Britain which belongs to the Pyrenean phylogroup, represented by individuals found in a few localities in the periphery of mainland Britain and in several British Isles. Compelling evidence indicate that pygmy shrews of the North-Central European lineage, as well as other small mammals, colonised Britain after the LGM and selectively displaced pygmy shrew populations of the Pyrenean lineage, which still remain in uplands and islands in the periphery to the north, west and south of Britain forming a 'Celtic fringe' (Searle et al. 2009), also observed in the sample analysed here.

### 2.4.6 Northern refugium of the Pyrenean phylogroup?

The significant population expansion signature, increasing population size shown in the BSPs and expansion times suggest that the Pyrenean lineage expanded during the Late Pleistocene. This lineage also may have had a refugium further north than traditionally recognized, separated from the southern refugium of the Iberian phylogroup and the northern refugia of the North-Central European phylogroup; it is restricted to Central, Western and South-Eastern France and North-Western Spain in continental Europe, but it is the only lineage found in Ireland and several islands off the west and north coasts of Britain. The region of the Dordogne in South-Western France was situated outside the LGM permafrost area and has temperate mammal fossil records dated to the end of the LGM; therefore, it has been suggested as another likely refugium north of the traditionally recognized southern refugia (Sommer and Nadachowski 2006). There, the Pyrenean lineage could have persisted and recolonised Western and Central France after the LGM, but SDM studies show that suitable climatic conditions during the LGM for *S. minutus* and other temperate small mammal species could have been more continuous and present further north (Fløjgaard et al. 2009, Vega et al. 2010b).

### 2.4.7 Origins of the Irish pygmy shrew

There are seven terrestrial small mammals in Ireland: *S. minutus*, *Apodemus sylvaticus*, *M. glareolus*, *Mus musculus*, *Rattus rattus*, *R. norvegicus* and *C. russula*, but only two have been present for a long time, *S. minutus* and *A. sylvaticus* (Martínková et al. 2007). How and when the pygmy shrew arrived and colonised Ireland is still under debate. If *S. minutus* arrived to Ireland on its own legs, then the colonisation would be concordant with the proposed migration route from Iberia and Western France to South-West Britain by an exposed coastline (Yalden 1982), but

this would not explain why other small mammal species did not arrive to Ireland following the same route. A possibility is that cold-tolerant small mammals survived close to the ice sheet at or before the LGM. For example, it was shown that the stoat *Mustela erminea* was present in Ireland before the LGM and spread throughout Ireland and Britain as glaciers retreated (Martínková et al. 2007). However, the pygmy shrew does not reach as far north as the stoat and it appears to be not as cold-tolerant, but a similar scenario cannot be entirely ruled out because the estimated population expansion times pre-date the LGM (although this could be related to the expansion signature found in the continental Pyrenean sample and there is a possibility of error in the molecular clock).

Alternatively, considerable phylogeographic and population genetics evidence for *S. minutus* has been gained favouring human-mediated introductions into Ireland and other islands off Britain from South-Western Europe during the Neolithic period (ca. 4 –7 KYA): Mascheretti et al. (2003) showed sequence similarity between Andorran and Irish *S. minutus*; and McDevitt et al. (2009) provided strong evidence based on microsatellite data for a population expansion event for the pygmy shrew in Ireland 6.3 KYA, in the Neolithic, but could not provide a source area. In this study, the central position of samples from Navarra, Spain in the MJ network and the overall similarity of Irish, Pyrenean and Western French samples support the hypothesis that *S. minutus* colonised Ireland from Northern Spain, not from Britain, its closest landmass, or Northern France. The low genetic variability in Ireland in comparison to that in Britain found in this and other studies (McDevitt et al. 2009) also supports a recent human-mediated colonisation, followed by a population expansion event. Nonetheless, there is a clear genetic similarity of Irish, Orkney, Celtic fringe and continental haplotypes, which could potentially indicate a colonisation route from South-Western Britain. Two issues yet need to be resolved: 1) the timing of colonisation, based on Rogers' (1995) $\tau$, does not concur

with a Neolithic colonisation; however, the mutation rate used here might not be suitable over short time frames (Ho et al. 2005, Ho and Larson 2006, but see Bandelt 2008), and 2) it still needs to be explained how a small mammal with high metabolic requirements such as the pygmy shrew survived a long trip and why other shrews, frequent in South-Western Europe, were not introduced the same way.

Further insights into the colonisation, migrations and morphological differences among Pyrenean, Irish, Orkney and the Celtic fringe samples will be discussed in Chapter 4.

### 2.4.8 Other considerations

The phylogeographic analyses presented here generally support the subspecific status of the genetic lineages described by Hutterer (1990), with distinct lineages in the Balkans, Iberia, Southern Italy and Northern-Central Europe, which roughly correspond to *S. m. gymnurus*, *S. m. carpetanus*, *S. m. lucanius* and *S. m. minutus*, respectively. However, there are no recent morphological analyses of *S. minutus* covering the full European range and there have not been comparisons between phylogeographic and morphometric findings to support this. As far as I know, there has only been one study that explored the morphological variation of *S. minutus* and compared it with molecular findings, but it covered a small part of the range and focused on the morphological and genetic variation within Italy and Northern-Central Europe (Vega et al. 2010a); the authors used a geometric morphometric technique coupled with tree-based phylogeographic analyses and showed that *S. minutus* from Southern Italy is genetically and morphologically distinct from northern populations, in agreement with the subspecific status of *S. m. lucanius*; however, the morphological and genetic differentiation in other regions was not analysed.

In the following chapter, I implement a geometric morphometrics approach and couple it with the phylogeographic findings obtained here to further explore the variation of *S. minutus* throughout its European range.

# Chapter 3

# Morphological diversity of *Sorex minutus* across Europe

## 3.1 Introduction

Morphometrics is the quantitative and comparative study of the variation and change of organismal form (Rohlf 1990). In biological studies, the analysis of morphological variation is of special concern for the understanding of ecological and evolutionary patterns and processes, and it is also important because organismal taxonomy is based on morphological traits. Several factors that cause morphological variation have been identified, including disease or injury, ontogenetic development, inbreeding, adaptation to local environments and long-term evolutionary diversification (Zelditch et al. 2004).

In the 1980s, the field of morphometrics was revolutionized (Rohlf and Marcus 1993, Adams et al. 2004) by the invention of coordinate-based methods, the development of the statistical theory of shape and its precise definition, and the computational use of transformation grids to visualize shape variation (Mitteroecker and Gunz 2009). This new morphometric approach was termed 'geometric morphometrics', which is based on the analysis of form (the shape and size of any object) based on Cartesian landmark coordinates, and the geometry of the landmark coordinates are preserved throughout the analysis (Zelditch et al. 2004, Mitteroecker and Gunz 2009).

The methods employed by geometric morphometrics originated from the fusion of geometric (or location) information with data on biological homology, combining geometry with biology (Bookstein 1982). Geometric morphometrics relies on the selection of landmarks, which are discrete anatomical points that should be

present in all specimens in the study, i.e., biologically homologous anatomical loci, also defined as points of correspondence on each object that match between and within populations (Dryden and Mardia 1998). To describe morphological variation properly, the selected landmarks must capture the form (size and shape) of the structure under study (Zelditch et al. 2004). Another characteristic of geometric morphometrics is the precise definition of shape and size, which in total represent the form of an object. Shape is defined as all the geometrical information that remains when scale, location and rotational effects are filtered out of an object (Kendall 1977). Size is defined by the scale of an object, which is mathematically independent of shape because it is filtered out according to Kendall's (1977) definition, and is calculated as the Centroid Size (CS): the square root of the sum of squared distances between each landmark and the centroid (or geometric centre of the object) (Bookstein 1991).

Because most morphological traits are under selection, the study of morphological variation can shed light on the impact of selection in adaptation and speciation processes; however, morphological variation is better understood when comparing it with results obtained from other methods, such as genetic, environmental, physiological or behavioral (Garnier et al. 2005). For example, several studies have used geometric morphometric techniques coupled with phylogenetic data and/or environmental correlates to provide a thorough characterization of species and population divergence with implications for the systematics, taxonomy and conservation of African murids (Fadda and Corti 2001), marmots (Cardini and O'Higgins 2004, Caumul and Polly 2005), spiny rats (Monteiro and dos Reis 2005) and ground squirrels (Gündüz et al. 2007), to evaluate the phylogenetic recoverability of skulls, mandibles and molars for the identification of karyotypic groups of fossils and infer divergence times of common shrew karyotypic races (Polly 2001, 2007), and to assess the correlation between

environmental components and geographical variation of vervet monkeys (Cardini et al. 2007) and punaré rats (Monteiro et al. 2003).

The wide geographic distribution of *Sorex minutus* and the amount of phylogeographic information available for this species makes it ideal for the study of intraspecific morphological variation. In this study I explored the following questions: 1) What is the morphological diversity of *S. minutus* and how is it distributed across Europe? 2) Do mandibles and skulls show the same or distinct patterns of morphological variation? 3) Can a latitudinal and/or longitudinal pattern for size and shape be identified? 4) Is there an allometric effect of size on shape? 5) To what extent does the morphological diversity resemble the phylogeographic structure detected previously with molecular markers? 6) To what extent does the morphological diversity resemble the taxonomical classification of *S. minutus* subspecies? To answer these questions, I employed a geometric morphometrics approach on mandible and skull samples of *S. minutus* obtained throughout its European range, and made use of the phylogeographic information obtained in the previous chapter. As far as I am aware, this is the first study that has analysed the morphological variation of *S. minutus* across its full European range based on a geometric morphometrics approach.

## 3.2 Materials and methods

3.2.1 Collection and preparation of samples

I acquired *S. minutus* and *S. volnuchini* skull and mandible samples during fieldwork and from museum or private collections. In total, I analysed 576 (left and right) mandibles and 385 skulls from Europe, including eight mandibles and skulls of *S. volnuchini* from the Anatolian peninsula, which were used as an outgroup where appropriate (Fig. 3.1A, B, Appendix 4). The sex was known for 254 mandible samples, of which 127 were males and 127 were females. For the skull sample,

there were 133 known males and 126 known females. Skulls and mandibles from specimens obtained during fieldwork were cleaned using a Papain method (White 2007) or using *Dermestes* sp. beetles. Some of the samples (mainly those from France) were obtained from owl pellets and were carefully cleaned using a small brush.
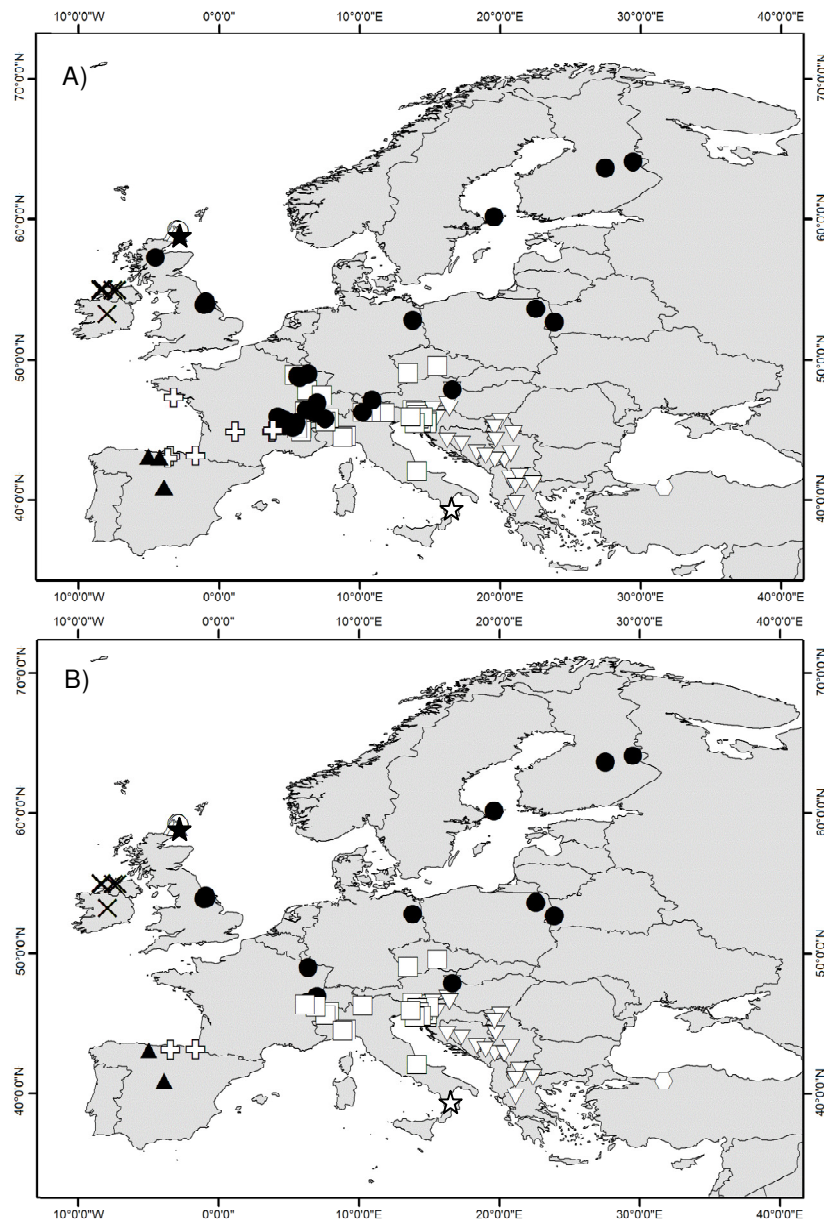


**Figure 3.1.** Samples of *Sorex minutus* for morphological analysis. A) Mandible dataset. B) Skull dataset. The symbols correspond to the different lineages detected previously with mtDNA. ● North-Central European, □ North-Central Italian, ▽ Balkan, ☆ South Italian, ▲ Iberian, ✚ Pyrenean (continental and Belle Île, France), △ Orkney Mainland, ○ Westray, ★ South Ronaldsay, X Ireland, ⬡ *S. volnuchini* (outgroup).

3.2.2 Digitization of mandibles and skulls

Photographic images of the external side of left and right mandibles and the ventral side of skulls were taken using a Canon EOS 350D camera (8 million pixels resolution) equipped with a macro lens and extension tubes at a fixed distance of 33 cm (for mandibles) and 34 cm (for skulls). Mandibles were placed on top of an inverted glass Petri dish (to avoid shadows surrounding the mandible) under the camera lens. Skull samples were placed on a purpose-built polystyrene and plasticine cradle leaving the ventral side parallel to the lens, judged by eye (Jones 2009). In each photograph, a small piece of graph paper was included as a scale and the sample was placed in the middle of the image area to avoid parallax.

Morphological analyses were carried out using the 'tps-Series' software (by FJ Rohlf, available at http://life.bio.sunysb.edu/morph/). Eighteen landmarks were placed on the external side of each mandible and 19 landmarks were placed on the right and left sides of the ventral cranium using tpsDig2 (Fig. 3.2A, B). The sample images were randomly ordered in tpsUtil before digitizing the landmarks to avoid any bias by repeatedly placing landmarks on samples from the same locality. The selected landmarks provided a comprehensive sampling of the morphology of the biological structures under study (Zelditch et al. 2004).
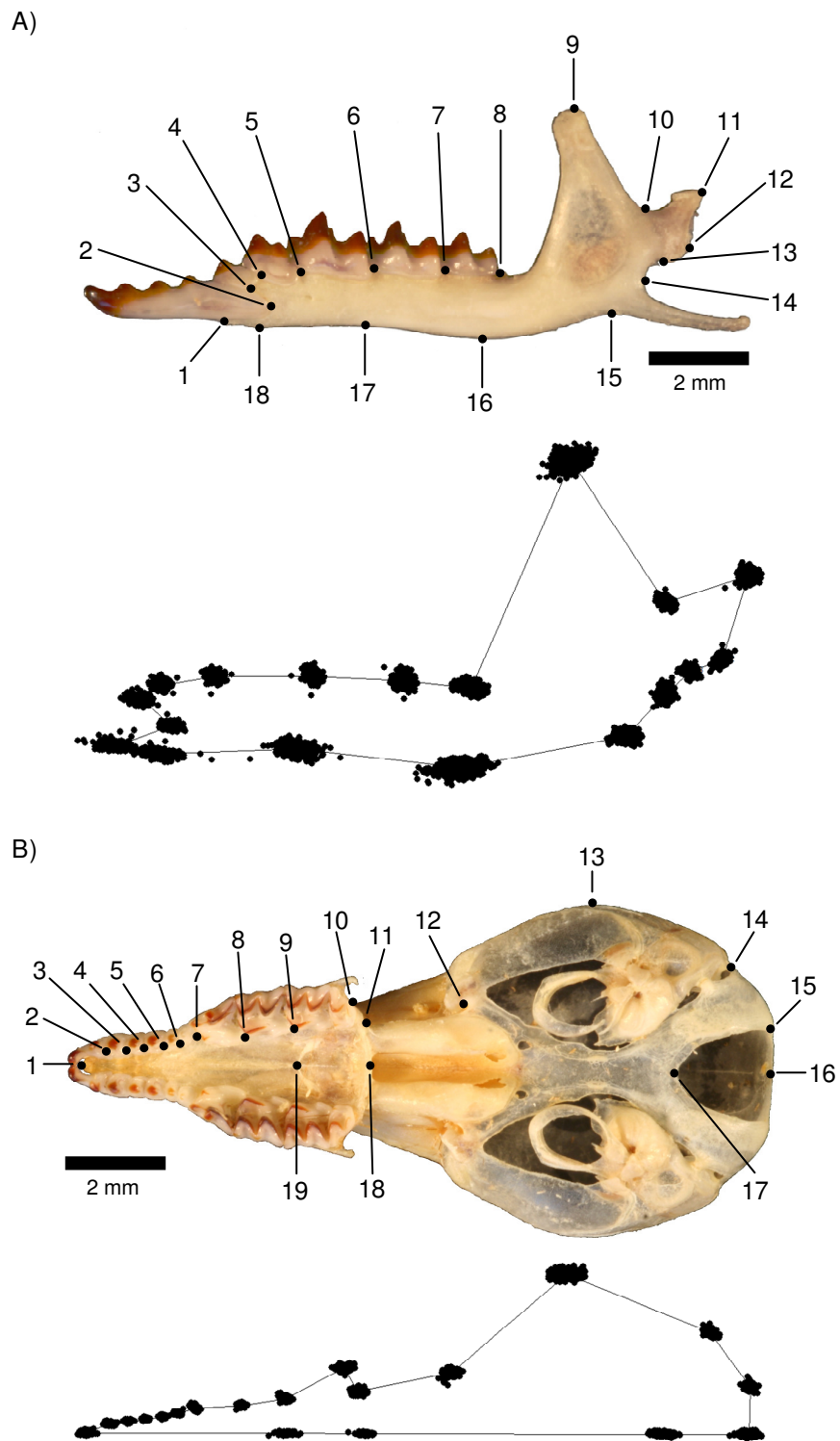
**Figure 3.2.** Landmarks placed on A) mandibles and B) skulls of *Sorex minutus*. For simplicity, only the left sides are shown. The diagrams show the observed dispersion of landmarks for all the aligned specimens.

3.2.3 Statistical analysis of landmark coordinates data

The statistical analyses were performed with SPSS/PASW Statistics version 17 (IBM), PAST version 1.97 (Hammer et al. 2001) or directly with the tps-Series software.

The size of each mandible and skull was estimated using the CS, obtained by tpsRelw and was transformed with natural logarithms. The landmark configurations were scaled to unit CS, translated and rotated using the Generalised Procrustes Analysis (GPA), which minimizes the differences between configurations by minimizing the Procrustes distances between corresponding landmarks and gives the average landmark configuration (Rohlf and Slice 1990). The Procrustes distance is the shortest distance between corresponding landmarks, and it is computed as the square root of the sum of squared distances between two centred, scaled and optimally rotated configurations of landmarks (Zelditch et al. 2004).

The Procrustes distance is an angle expressed in radians, which gives the geodesic distance between two points in Kendall's shape space, and it corresponds to the distance along the surface of a sphere (Kendall 1977). Because all shapes are in non-Euclidean shape space (i.e., lie in Kendall's shape space), the shortest distance between two points is a curve (Zelditch et al. 2004). To perform statistical analyses, the Procrustes distances have to be approximated to an Euclidean space using, in this case, an orthogonal projection. The average configuration is the point of tangency between shape space and Euclidean shape space. Then, I used tpsSmall to test if shape variation in the mandible and skull samples was small enough to allow an approximation to Euclidean tangent space. A strong positive correlation ($R = 0.9999$) was found between tangent space distances and Procrustes distances among samples; therefore, the orthogonal tangent space approximation for shape analysis was validated.

A weight matrix (*W*) incorporating Uniform (n = 2 for mandibles and skulls) and non-Uniform (n = 30 for mandibles and n = 32 for skulls) components was extracted using the GPA (Bookstein 1996). Both components were interpreted and pooled together as shape variables (n = 32 for mandibles and n = 34 for skulls). The Uniform components of shape deformation do not require bending energy to shift from the average configuration to the observed shape, including transformations such as translation, rotation and scaling, which do not alter shape, and compression, dilation and shear, which do alter shape. The non-Uniform components do require bending energy to alter shape, have a location and a spatial extent, and describe the patterns of relative landmark displacements in relation to the average configuration (Zelditch et al. 2004).

First, to check for landmark placement repeatability, I photographed and digitized 23 mandibles and 28 skulls five times each, to give a total of 115 and 140 photographs, respectively. Digital images were randomized with tpsUtil and the appropriate landmarks were placed on each mandible and skull in tpsDig2. The CS values of mandibles and skulls among photographs were subjected to Analysis of Variance (ANOVA) and the *W* matrices among photographs were analysed with Multivariate Analysis of Variance (MANOVA) (only the landmark configurations on left mandibles and the left side of skulls were analysed). Second, the X coordinates of right mandibles and the Y coordinates of the right side of skulls were multiplied by −1 to create mirror images for both sides of the bilateral structures. Then, left and right mandibles and the left and right sides of skulls were compared by ANOVA for CS, and by MANOVA or non-parametric MANOVA (Anderson 2001) for *W* to check for asymmetry. Third, male and female samples were compared by ANOVA for CS, and by MANOVA or non-parametric MANOVA for *W* to check for sexual dimorphism.

The pattern of shape variation related to mandible and skull size change (allometry) was compared for the *S. minutus* data set (excluding the outgroup) using tpsRegr. The optimal measure of allometry is the regression of shape on the logarithm of CS (Mitteroecker and Gunz 2009). Multiple regressions of shape variables on latitude and longitude for the *S. minutus* data set (excluding the outgroup) were also performed with tpsRegr to estimate and visualise their effects on shape.

Because of the relatively small number of individuals from each location, and for the purpose of this study, I pooled the samples into 11 mutually exclusive groups according to their cytochrome (cyt) *b* phylogroup membership, designated as Iberian, Balkan, North-Central Italian, North-Central European, South Italian, Pyrenean (including the samples from the continent and Belle Île, France) and outgroup (*S. volnuchini* samples), and to their island origin, designated as Ireland, Orkney Mainland, Westray and South Ronaldsay groups. Normality tests were performed for the CS of all groups. The mandible CS distributions of Balkan, Iberian and South Italian groups, and skull CS distributions of Iberian, Orkney Mainland and South Italian groups were significantly different from a normal distribution, but for the purpose of this study were not excluded from the analysis; quantile-quantile plots only showed small deviations of the observed CS from the expected normal distributions (data not presented).

Mandible and skull size differences among groups were evaluated by ANOVAs and visualised with box plots. I performed a Tukey–Kramer post-hoc test on CS as it allows for unequal sample size (Sokal and Rohlf 1995). Levene's test was performed to detect heteroscedasticity (non-equality of error of variances across groups). Although ANOVAs are relatively robust to heteroscedasticity (Zelditch et al. 2004), if Levene's test gave significant results, I also performed the conservative non-parametric Kruskal-Wallis test for the analysis of CS variance and

a Tamhane non-parametric post-hoc test, both of which do not assume equal variances. A linear regression was performed to assess how group mean CS for mandibles and skulls varied with latitude and longitude (in decimals) treated as independent variables.

Mandible and skull shape differences among groups were evaluated with MANOVAs on $W$ (i.e., shape variables), followed by Hotelling $T^2$ tests for multivariate comparisons performed in PAST. Levene's tests were also performed for each shape variable among groups in SPSS and, if significant deviations from the assumption of homogeneity of variances were found, non-parametric MANOVAs among groups were performed in PAST.

Ordination of the samples was carried out through Principal Component Analysis, i.e., Relative Warp (RW) Analysis (Rohlf 1993), on the shape variables using tpsRelw for the mandible and skull data sets. RW Analysis was also conducted on the average (consensus) configurations of the 11 groups for the mandible and skull data sets. RW Analysis replaces the original shape variables (the $W$ matrix) with Principal Components, which are linear combinations of the original variables but independent from each other (Zelditch et al. 2004). Shape changes in the RW space were visualised as thin-plate spline (TPS) transformation grids (Bookstein 1989) obtained with tpsSpline. The TPS is a continuous interpolating function that extends between the landmarks and describes the shape variation as a deformation from the average configuration, which can be visualized as transformation grids (Zelditch et al. 2004).

The Procrustes distances among the average configurations of the groups, for the mandible and skull data sets, were then computed by the program tpsSmall and entered into PAST to produce a distance tree using the Neighbor-Joining (NJ) method (discussed in Chapter 2) to evaluate the morphological relationships. The support of the branches was assessed using 10,000 bootstraps.

The geographic midpoints for the 11 groups, using the coordinates of the sampling localities, were calculated with the Geographic Midpoint Calculator (available at http://www.geomidpoint.com/). The geographic midpoints were then used to obtain the pairwise geographic distances among groups with the Geographic Distance Matrix Calculator version 1.2.3 (by PJ Ersts, available at http://biodiversityinformatics.amnh.org/open_source/gdmg). A Mantel test was performed in PAST to examine the correlation between the pairwise Procrustes distances and the pairwise geographic distances among groups. Also, Mantel tests were performed between pairwise Procrustes distances and pairwise genetic differentiation ($F_{ST}$) and genetic divergence (Da) values of the cyt *b* phylogroups (obtained in Chapter 2). The significance of the test was obtained through a permutation procedure (10,000 bootstraps) used to create an empirical null distribution. The overall shape diversity for mandibles and skulls was visualised as TPS deformations and plotted next to the NJ trees of Procrustes distances among the groups.

A Discriminant Analysis (DA) was performed to predict group membership and to differentiate among the groups by a Canonical Variates Analysis (CVA), using the shape variables as predictors for mandible and skull data sets. The first two functions were used to graph the samples separated by group membership. The DA builds a model to predict the group membership of samples using linear combinations of the predictor variables that best discriminate between the groups. The functions are generated from a selection of samples for which group membership is known and then applied to samples with unknown group membership. I used a cross-validation method in SPSS, repeatedly treating n – 1 out of n samples as the 'validating' dataset to determine the discrimination rule and using the rule to classify the one observation left out. I also implemented a pre-validation method in SPSS to predict group membership, selecting those samples

with 'known' group membership for 'training' the model (i.e., using the samples for which I could be absolutely confident about group membership, such as island samples or morphological samples with genetic data) and leaving other samples aside, then applying the trained model to predict group membership of the unused samples.

**3.3 Results**

3.3.1 Landmark placing error

There were no significant differences with respect to CS or $W$ among the five repeated photographs of mandibles ($F_{4, 110} = 0.0050$, $p > 0.05$; Wilks' $\lambda = 0.5117$, $F_{128, 316.9} = 0.4499$, $p > 0.05$) and skulls ($F_{4, 135} = 0.0700$, $p > 0.05$; Wilks' $\lambda = 0.5666$, $F_{136, 408.7} = 0.4610$, $p > 0.05$); therefore, landmarks were placed accurately in each sample and the variation found in this study was not related to the misplacement of landmarks.

3.3.2 Asymmetry and sex differences

There were no significant differences with respect to CS between left and right mandibles ($F_{1, 1150} = 1.7770$, $p = 0.1830$) or the right and left sides of skulls ($F_{1, 768} = 0.1490$, $p = 0.6990$). However, there were significant shape differences for the same comparisons of mandibles (Wilks' $\lambda = 0.6197$, $F_{32, 1119} = 21.4600$, $p < 0.0001$) and skulls (Wilks' $\lambda = 0.4434$, $F_{34, 735} = 27.1400$, $p < 0.0001$). The assumption of homogeneity of variance for $W$ of mandibles and skulls was violated, but non-parametric MANOVAs also revealed significant differences between left and right mandibles ($F_{32, 1119} = 21.4600$, $p < 0.0001$) and the right and left sides of skulls ($F_{34, 735}$, $p < 0.0001$). To avoid the effects of asymmetry, I averaged both sides of mandibles and skulls in all subsequent analyses.

There were no significant differences between male and female mandibles for CS ($F_{1, 148} = 0.4840$, $p = 0.4870$) and $W$ (Wilks' $\lambda = 0.8123$, $F_{34, 117} = 0.8447$, $p = 0.7030$; non-parametric MANOVA $F_{34, 117} = 1.4250$, $p = 0.1710$). Similarly, there were no significant differences between male and female skulls for CS ($F_{1, 516} = 0.0320$, $p = 0.8580$) and $W$ (Wilks' $\lambda = 0.9220$, $F_{34, 483} = 1.206$, $p = 0.2010$; non-parametric MANOVA $F_{34, 483} = 1.1290$, $p = 0.3105$). All subsequent analyses on mandibles and skulls were performed pooling all samples irrespective of sex.

3.3.3 Morphometric analysis of mandibles

Mandible CS decreased notably with increasing latitude (Table 3.1), as shown by box plots (Fig. 3.3). The South Italian, Iberian and Balkan groups, as well as the outgroup, had the biggest mandibles with quartile boxes well above the sample average. The median CS of Pyrenean and North-Central Italian mandibles were also bigger than the sample average, while the other groups (North-Central European, Ireland, Orkney Mainland, South Ronaldsay and Westray) had median values below the sample average. Nevertheless, the group ranges were very widespread and overlapping, especially for the Balkan, Pyrenean and North-Central Italian groups.

**Table 3.1.** Centroid size statistics for *Sorex minutus* mandibles.

| Phylogroup | n | Mean | Median | SD | SE |
|---|---|---|---|---|---|
| Balkan | 51 | 2.4065 | 2.4097 | 0.0303 | 0.0042 |
| Iberian | 13 | 2.3997 | 2.4028 | 0.0220 | 0.0061 |
| Ireland | 63 | 2.3486 | 2.3505 | 0.0232 | 0.0029 |
| North-Central European | 150 | 2.3359 | 2.3388 | 0.0279 | 0.0023 |
| North-Central Italian | 79 | 2.3803 | 2.3798 | 0.0279 | 0.0023 |
| Orkney Mainland | 52 | 2.3575 | 2.3606 | 0.0231 | 0.0032 |
| Orkney South Ronaldsay | 40 | 2.3390 | 2.3392 | 0.0141 | 0.0022 |
| Orkney Westray | 40 | 2.3590 | 2.3567 | 0.0231 | 0.0032 |
| Pyrenean | 77 | 2.3719 | 2.3730 | 0.0353 | 0.0040 |
| South Italian | 3 | 2.4053 | 2.4065 | 0.0289 | 0.0167 |
| Outgroup | 8 | 2.4165 | 2.4114 | 0.0214 | 0.0076 |
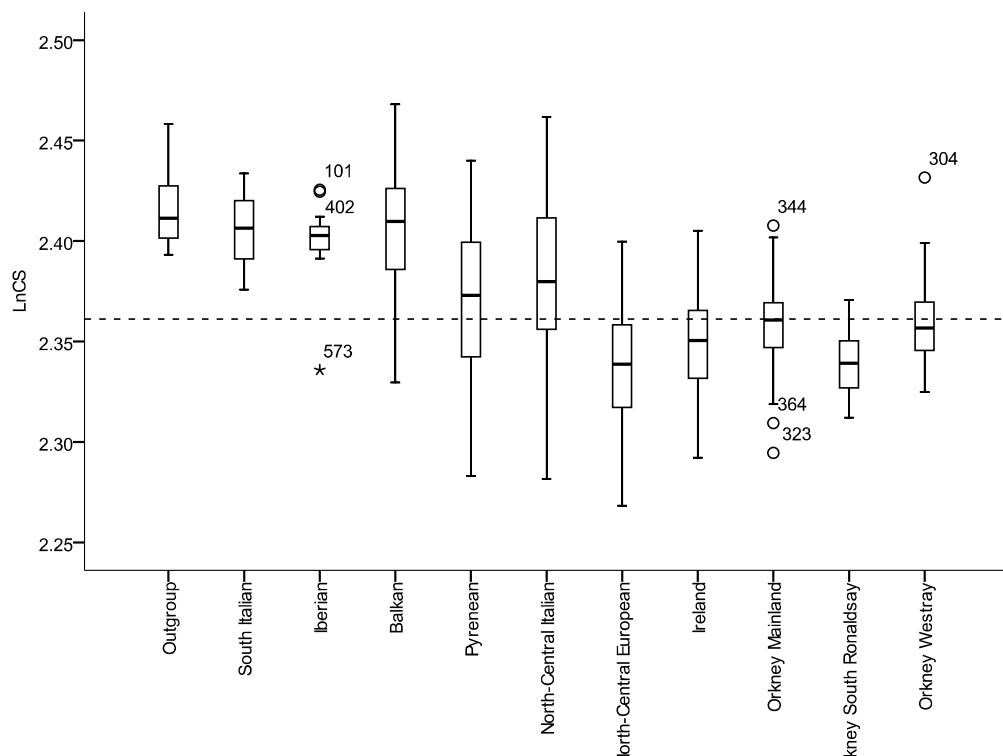
**Figure 3.3.** Box plots of mandible Centroid Size (CS; transformed with natural logarithms) of *Sorex minutus*. There is a decrease of CS from Southern (left) to Northern (right) latitudes. The horizontal dashed line corresponds to the sample mean. The outgroup (*S. volnuchini*) is shown for purposes of comparison.

The ANOVA revealed that there were significant CS differences among the *S. minutus* groups ($F_{9, 558} = 38.8850$, $p < 0.0001$); however, because the assumption of homogeneity of variances was violated (Levene's statistic$_{9, 558} = 8.5210$, $p < 0.0001$), a non-parametric test was performed and it was also significant ($H_9 = 205.5730$, $p < 0.0001$). Parametric (Tukey–Kramer) and non-parametric (Tamhane T2) post-hoc tests showed that nearly 57% of the pairwise comparisons were significantly different (Appendix 5). In the non-parametric post-hoc tests, the South Italian, Iberian and Balkan groups were not significantly different from each other, having similar and large mandible size. Notably, the mandibles CS of the Pyrenean

and Ireland groups were significantly different, but Ireland was not significantly different from any of the Orkney islands, and only South Ronaldsay was significantly different from the Pyrenean group and the other two Orkney islands.

The linear regressions showed different and significant tendencies of CS with latitude and longitude (Fig. 3.4A, B); I found decreasing CS with increasing latitude, as expected from the box plots, but also an increasing CS with increasing longitude. There was a relatively stronger relationship and higher proportion of CS variation explained by latitude ($R_{Latitude} = -0.4580$, $R^2_{Latitude} = 0.2096$, $p < 0.0001$) than longitude ($R_{Longitude} = 0.2230$, $R^2_{Longitude} = 0.0499$, $p < 0.0001$).

There was a highly significant allometric effect of CS on mandible shape (Goodall's $F_{32, 18112} = 11.1718$, $p < 0.0001$), but it explained a small percentage of shape variation among individuals (about 2%). In larger mandibles, there was an apparent forward movement of the landmarks on the lower part of the mandible (landmarks 1 and 16 – 18), in relation to the landmarks between teeth alveoli (landmarks 3 – 8), and there was some variation in the position of the condyloid process (landmarks 10 – 14) in relation to other landmarks (Fig. 3.5A).

Latitude and longitude had highly significant effects on mandible shape (Latitude Goodall's $F_{32, 18112} = 30.5923$, $p < 0.0001$; Longitude Goodall's $F_{32, 18112} = 18.8495$, $p < 0.0001$) and explained a higher percentage of shape variation than CS variation (5.13% and 3.22%, respectively). In individuals from southern latitudes there was a relative forward shift of the coronoid process, plus a backward movement of landmarks 3 – 8 between teeth alveoli and forward movement of landmarks 1, 16 – 18 (Fig. 3.5B). Individuals in eastern longitudes also showed a relative forward shift of the coronoid process, backward movement of landmarks between teeth alveoli and forward movement of landmarks on the lower side of the mandible, but these changes were less obvious than with latitude (Fig. 3.5C).

A)

$R^2 = 0.2096$

B)

$R^2 = 0.0499$
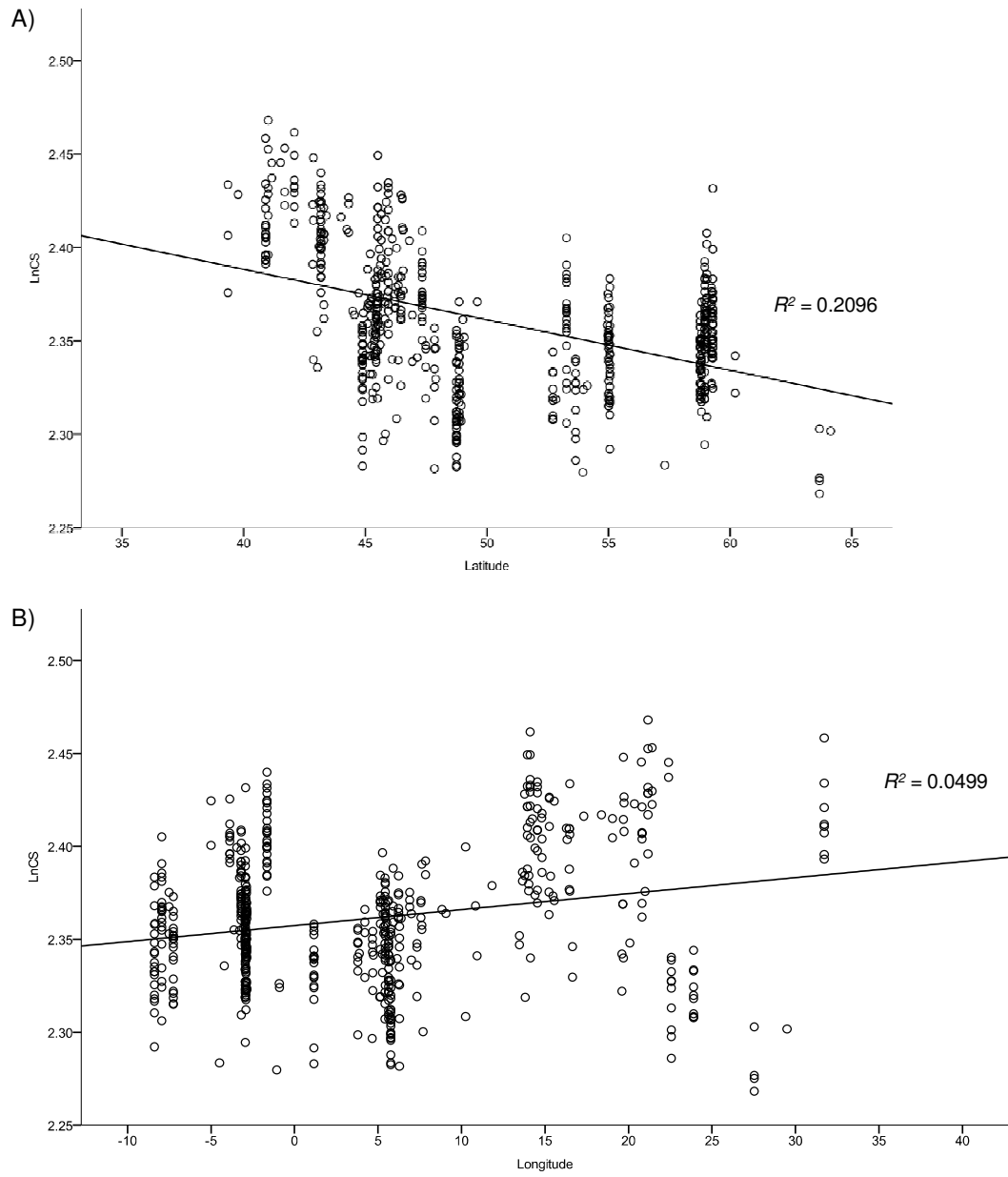
**Figure 3.4.** Linear regressions of Centroid Size (CS; natural log-transformed) on A) latitude and B) longitude for *Sorex minutus* mandible samples.

A)



Shape vs. CS
3 X

Maximum CS                                         Minimum CS

B)



Shape vs. Latitude
3 X

South                                              North

C)



Shape vs. Longitude
3 X

West                                               East
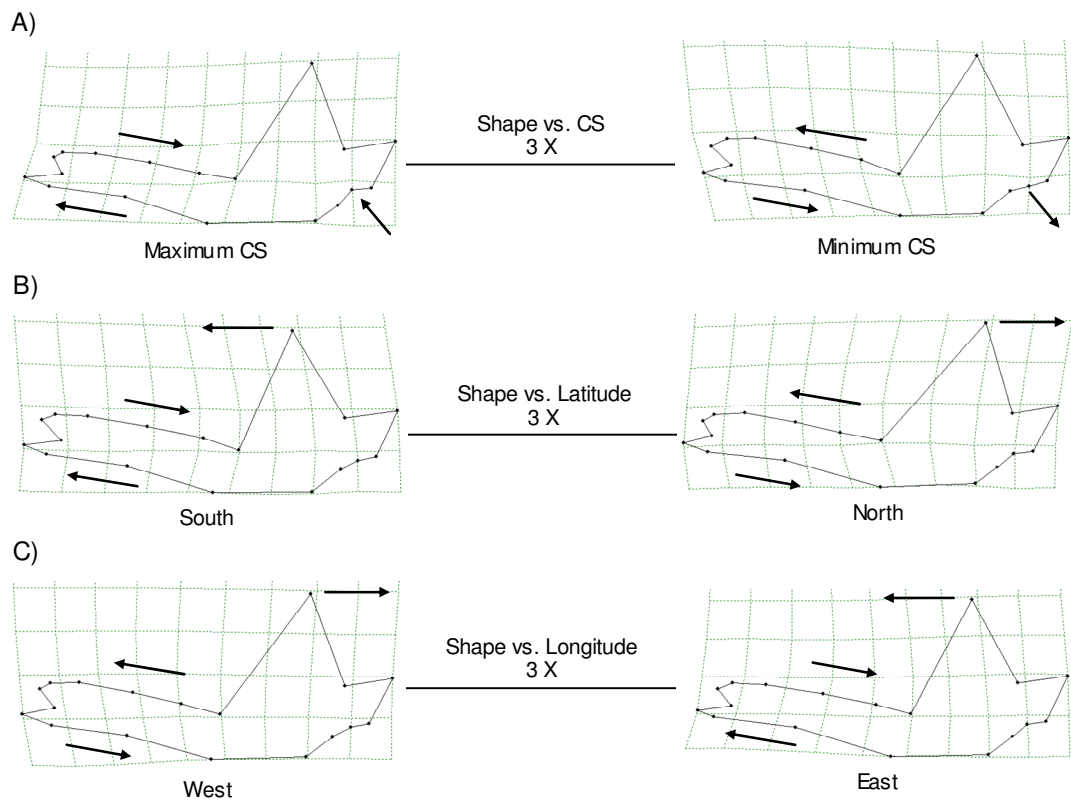
**Figure 3.5.** Multiple regressions of shape on A) Centroid Size (CS), B) latitude and C) longitude for *Sorex minutus* mandibles. Transformation grids were exaggerated 3 times for purposes of visualisation. Arrows help to evaluate the relative movements of landmarks.

The RW Analysis showed that shape variation was not subdivided in any systematic way. There was a nearly uniform ordination of the samples around the average configuration (Fig. 3.6). The first three RWs explained cumulatively 48.26% of the total variance (20.21%, 14.52% and 13.53%, respectively). There was little shape deviation from the average configuration, but the landmark movements along RW1 seem to follow the latitudinal pattern described previously.
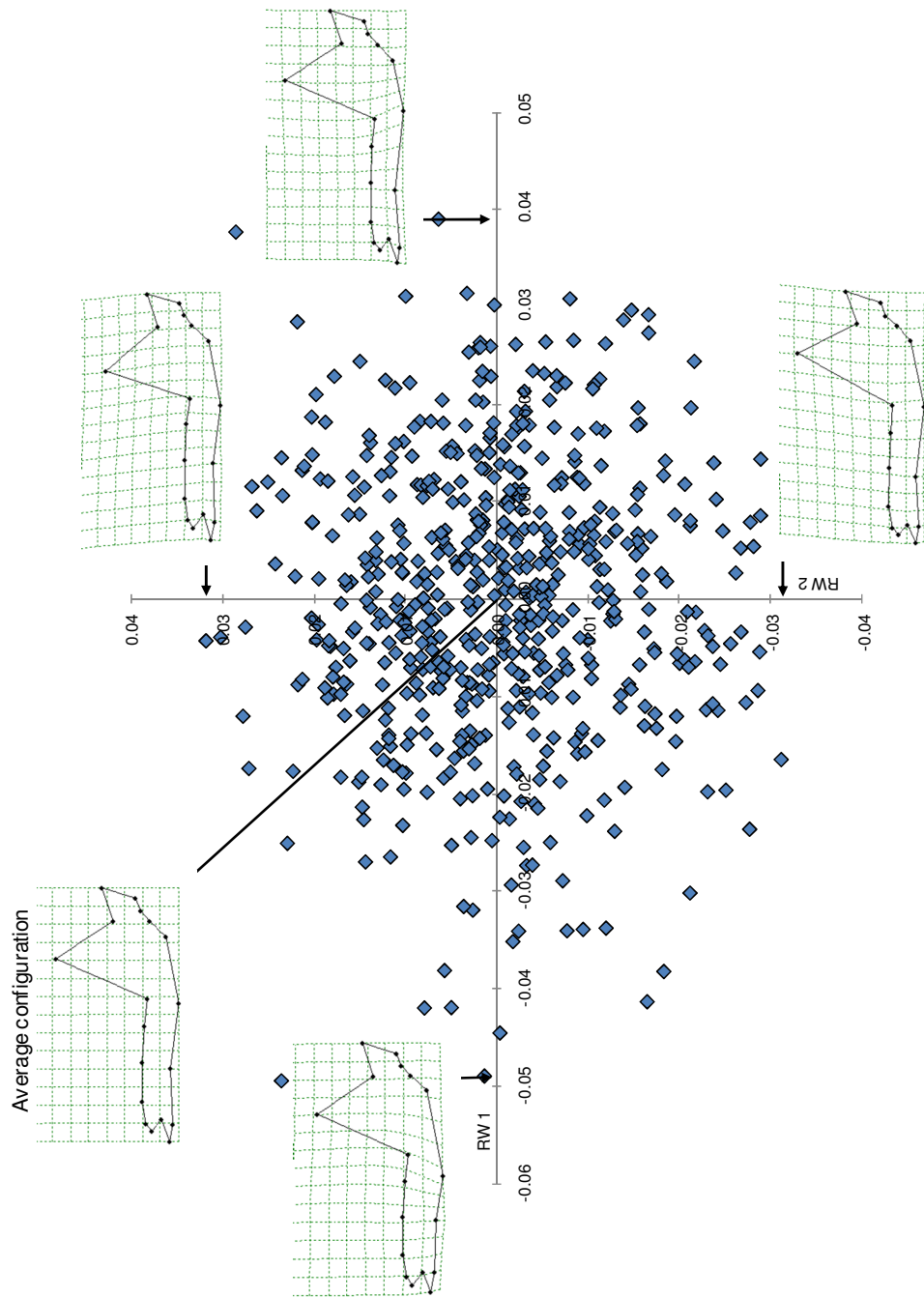
**Figure 3.6.** Relative Warp Analysis of shape variables of *Sorex minutus* mandibles. The maximum and minimum transformation grids on the first two Relative Warps (RW1 and RW2), and the average configuration are shown for purposes of visualisation.

The parametric and non-parametric MANOVAs, in contrast, revealed significant differences among groups (Wilks' $\lambda = 6.3 \times 10^{-5}$, $F_{320, 5136} = 28.02$, $p < 0.0001$; non-parametric MANOVA $F_{320, 5136} = 10$, $p < 0.0001$), and the post-hoc tests showed that nearly 70% of all pairwise comparisons were significant (Appendix 6). Most of the non-significant differences involved comparisons with the South Italian and Iberian groups, likely a reflection of the small sample size. Including the CS as a covariate of shape variables or including the outgroup in the MANOVA made no differences to the outcome.

The discriminant function analysis allowed further exploration of the differences in mandible shape among groups (separated into phylogroups and island groups, accordingly). This analysis correctly classified 67.7% of the individuals to their predefined group, but using a cross-validation method only 55.4% were properly classified, a matter of chance rather than clear shape differences among groups. Using a pre-validation method, training the classification method with those samples with 'known' group membership, the discriminant function analysis correctly classified 86.9% of selected individuals, and it decreased to 67.3% with the cross-validation method.

For purposes of visualisation, a scatter plot of the first two canonical variates is presented with group memberships and their group centroids (Fig. 3.7): the shape distribution of the North-Central European, Pyrenean, North-Central Italian, South Italian, Balkan and Iberian groups mostly overlapped; however, island samples were more easily discriminated from the rest, especially those from Westray, suggesting that there were shape differences between island and continental samples.
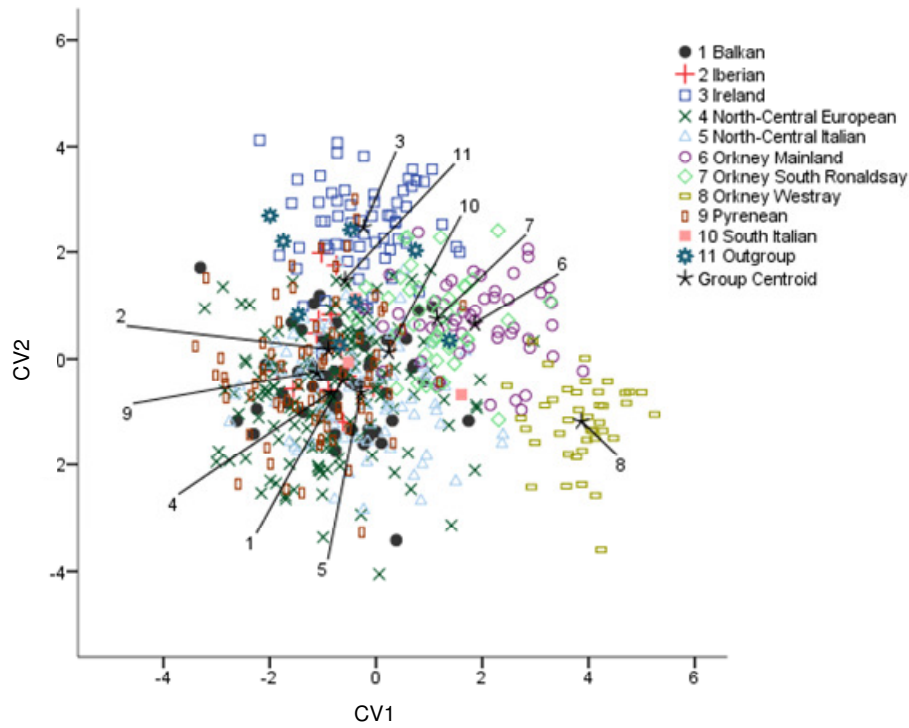
**Figure 3.7.** Canonical Variate Analysis of shape variables of *Sorex minutus* mandibles showing group differences. The first two canonical variates (CV) were used to describe variation among groups.

Similarly, the unrooted NJ tree of Procrustes distances among groups revealed that island samples clustered separately from continental ones, with the Orkney islands clustering with the South Italian group and the Outgroup, but separate from the Pyrenean, Iberian, Balkan, North-Central European and North-Central Italian groups (Fig. 3.8A). The South Italian, Westray and Outgroup groups were morphologically very distinct from any other groups, with large Procrustes distances, while the Orkney Mainland and South Ronaldsay groups clustered together (Fig. 3.8A). The rooted tree showed a slightly different topology, with Ireland closer to continental groups than to the Orkney islands, and the continental

groups clustering together (Fig. 3.8B). Rooting the tree still showed that Orkney Mainland and South Ronaldsay, and the North-Central Italian and North-Central European groups formed supported groups, and that the South Italian and Westray groups were the most morphologically divergent (Fig. 3.8B).

The transformation grids revealed that the three Orkney islands had notable backward shifts of the coronoid process in comparison to other groups, with Westray also showing pronounced variation in the frontal part of the mandible, whereas in the Iberian and Balkan groups the coronoid process moved slightly forward (Fig. 3.8A, B). The shape variation corresponded to the latitudinal shape change described previously. The South Italian group also had marked landmark variation and had a wider mandible (Fig. 3.8A, B); however, the relatively high pairwise Procrustes distances could be a result of small sample size.

There was no similarity between the NJ trees of Procrustes distances and the phylogenetic tree, and the amount of morphological variation did not seem to parallel the observed genetic distances (Fig. 3.8). Moreover, the Mantel tests revealed that there were significant correlations between Procrustes distances and geographic distances ($R = 0.3706$, $p = 0.0252$), but not with genetic differentiation ($F_{ST}$; $R = -0.0509$, $p = 0.6144$) or genetic divergence (Da; $R = 0.0645$, $p = 0.2610$) among groups. These results were consistent with the correlations of shape with latitude and longitude.

A)

Cyt b

0.005

76

75

97

60

83

73

53

0.005

North-Central Italian

North-Central European

Balkan

Orkney South Ronaldsay

Orkney Mainland

South Italian

Outgroup

Orkney Westray

Ireland

Pyrenean

Iberian

Sample average

(Continued)

92
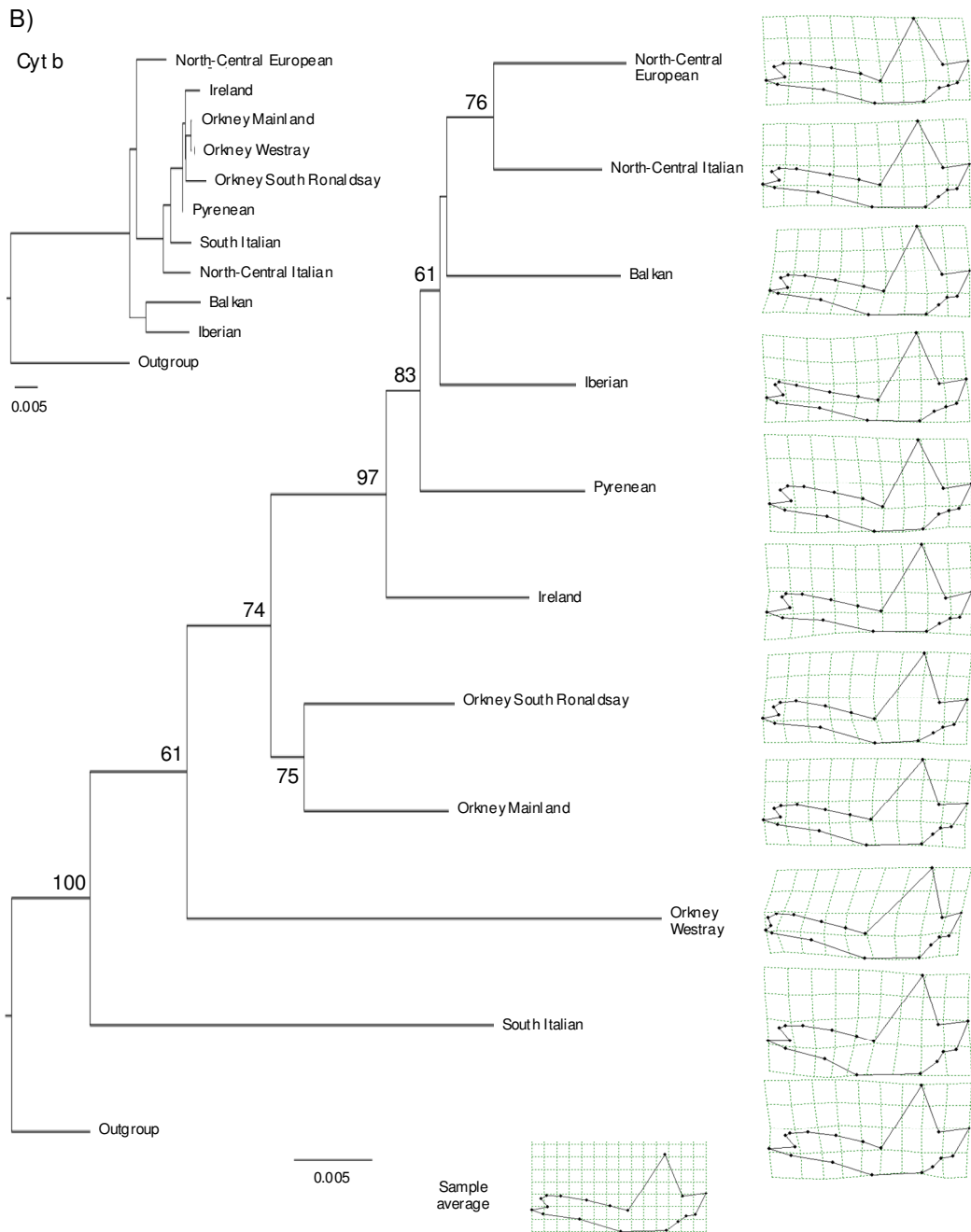
**Figure 3.8.** A) Unrooted and B) rooted Neighbor-Joining (NJ) trees constructed using mandible Procrustes distances among *Sorex minutus* groups. Branch support is shown as bootstrap values (10,000 replications). The grids on the right correspond to shape deformation per group from the average configuration. Inset: NJ trees based on genetic distances obtained in Chapter 2, shown for purposes of comparison.

3.3.4 Morphometric analysis of skulls

Skull CS showed less variation than mandible size among European regions, and had no apparent decreasing or increasing tendency with latitude (Table 3.2), as shown by the box plots (Fig. 3.9). The Iberian group had the biggest skulls, while the North-Central European had the smallest skulls on average but with a large variability. The median CS for the rest of the groups was similar to the sample average.

**Table 3.2.** Centroid size statistics for *Sorex minutus* skulls.

| Phylogroup | n | Mean | Median | SD | SE |
|---|---|---|---|---|---|
| Balkan | 46 | 3.0928 | 3.0949 | 0.0276 | 0.0041 |
| Iberian | 12 | 3.1270 | 3.1366 | 0.0354 | 0.0102 |
| Ireland | 60 | 3.0901 | 3.0897 | 0.0388 | 0.0058 |
| North-Central European | 45 | 3.0688 | 3.0654 | 0.0388 | 0.0058 |
| North-Central Italian | 57 | 3.0949 | 3.0943 | 0.0221 | 0.0029 |
| Orkney Mainland | 50 | 3.0983 | 3.0951 | 0.0280 | 0.0040 |
| Orkney South Ronaldsay | 37 | 3.0983 | 3.0997 | 0.0127 | 0.0021 |
| Orkney Westray | 39 | 3.1044 | 3.1048 | 0.0131 | 0.0021 |
| Pyrenean | 28 | 3.0947 | 3.0951 | 0.0174 | 0.0033 |
| South Italian | 3 | 3.0929 | 3.1002 | 0.0177 | 0.0102 |
| Outgroup | 8 | 3.0946 | 3.1012 | 0.0198 | 0.0070 |
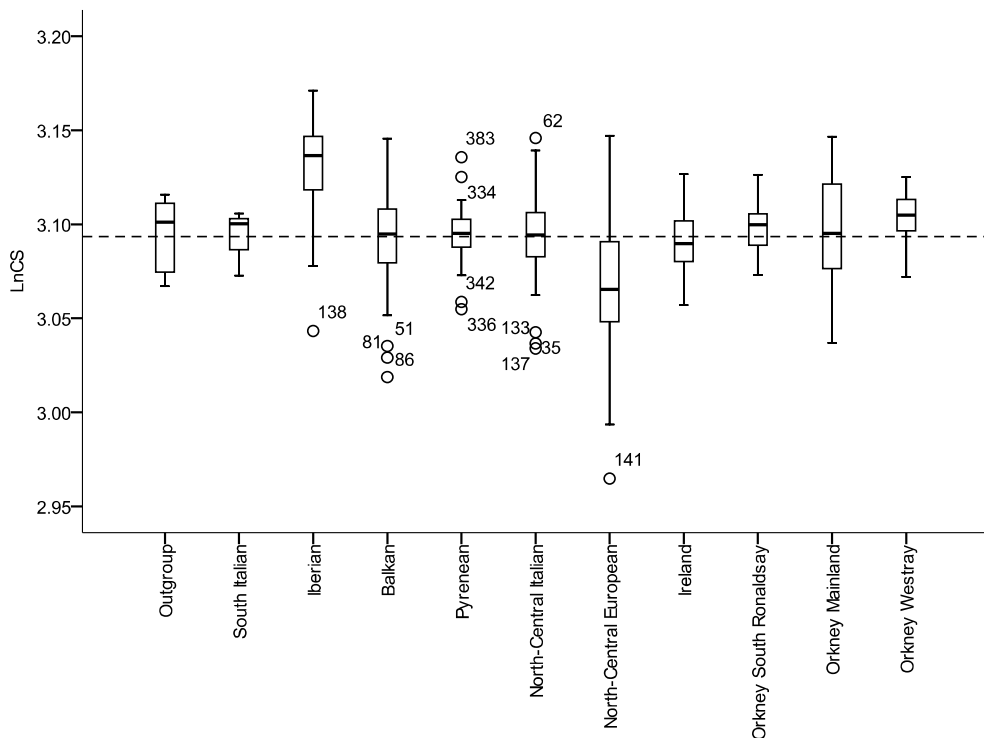
**Figure 3.9.** Box plots of skull Centroid Size (CS; natural log-transformed) of *Sorex minutus*. There is no apparent tendency from Southern (left) to Northern (right) latitudes. The horizontal dashed line corresponds to the sample mean. The outgroup (*S. volnuchini*) is shown for purposes of comparison.

The ANOVA on skull CS also revealed, as with mandibles, that there were significant differences among groups ($F_{9, 367} = 9.1040$, $p < 0.0001$); however, the assumption of homogeneity of variances was violated (Levene's statistic$_{9, 367} = 7.9060$, $p < 0.0001$), but the non-parametric test was also significant ($H_9 = 53.1750$, $p < 0.0001$). Parametric (Tukey–Kramer) and non-parametric (Tamhane T2) post-hoc tests revealed that nearly 31% and 20% pairwise comparisons were significantly different, respectively, a lower number of significant differences with respect to CS of skulls than for mandibles. In the parametric post-hoc tests all the significant pairwise comparisons involved the Iberian and North-Central European groups, but in the non-parametric tests the significant values only involved the

North-Central European group, except for a significant value between Ireland and Westray groups (Appendix 7).

Skull CS showed a minor decreasing tendency with increasing latitude and longitude. The linear regression of skull CS on latitude (Fig. 3.10A) was significant but less strong than that previously found in mandibles ($R_{Latitude} = -0.1090$, $R^2_{Latitude}$ = 0.0120, $p$ = 0.0320). The linear regression of skull CS on longitude (Fig. 3.10B) was significant and stronger than latitude, and it was as strong as that previously found with mandibles but it was negatively correlated ($R_{Longitude} = -0.2220$, $R^2_{Longitude}$ = 0.0490, $p < 0.0001$).

There was a highly significant allometric effect of CS on skull shape (Goodall's $F_{34, 12750}$ = 4.7488, $p < 0.0001$), but it only explained a small percentage of shape variation among individuals (1.25 %), in a similar fashion as found for mandibles. In bigger skulls, there was an apparent outward movement of landmark 7 in relation to other landmarks between teeth alveoli (landmarks 2 – 6, 8 and 9) resulting in a wider separation of the upper premolars, there was an inward relative movement of landmark 2 resulting in narrower and more pointed snouts, and there were backward movements of landmarks 18 and 19 resulting in the lengthening of the maxilla and pre-maxilla (Fig. 3.11A).
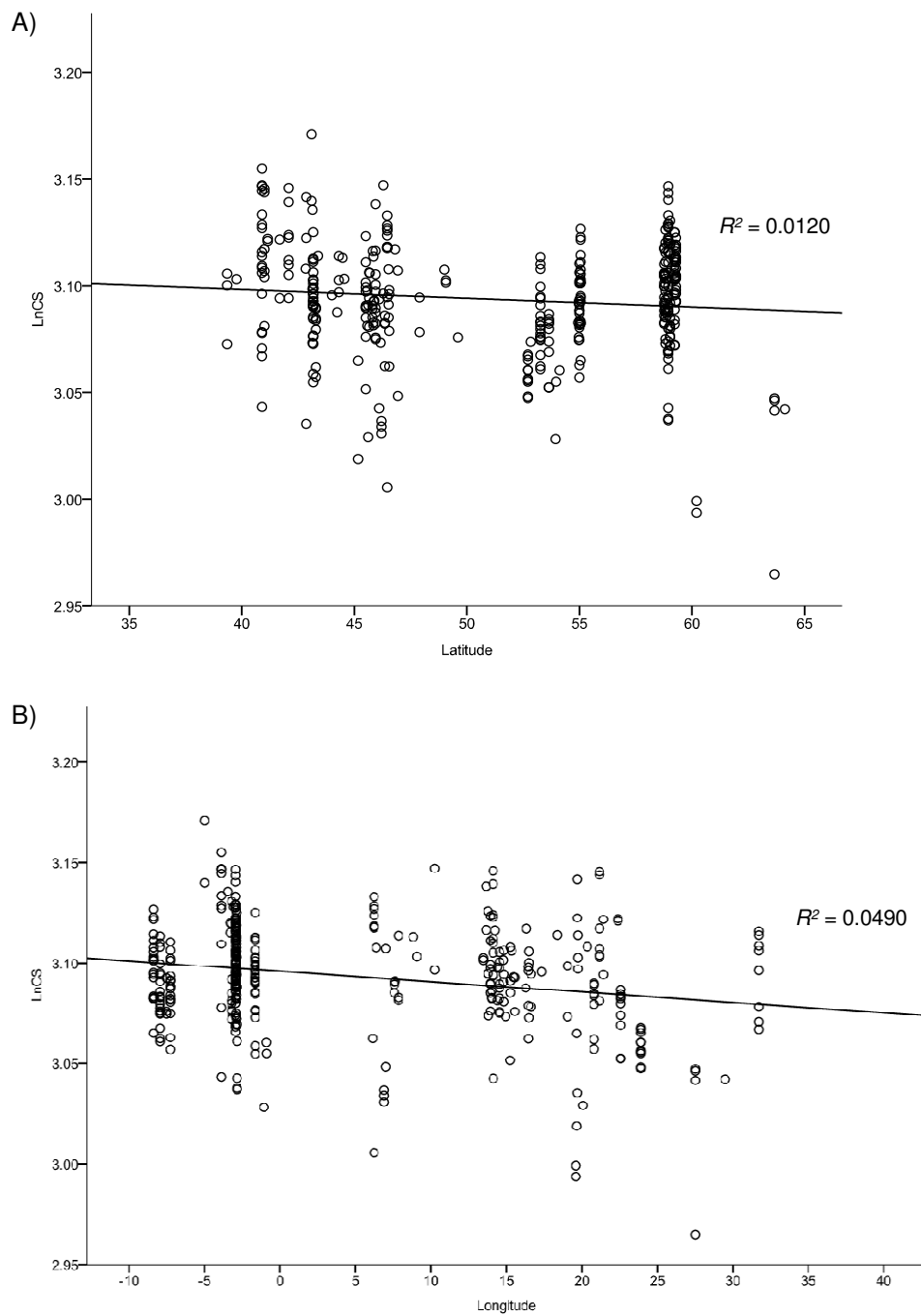
**Figure 3.10.** Linear regressions of Centroid Size (CS; transformed with natural logarithms) on A) latitude and B) longitude for *Sorex minutus* skull samples.
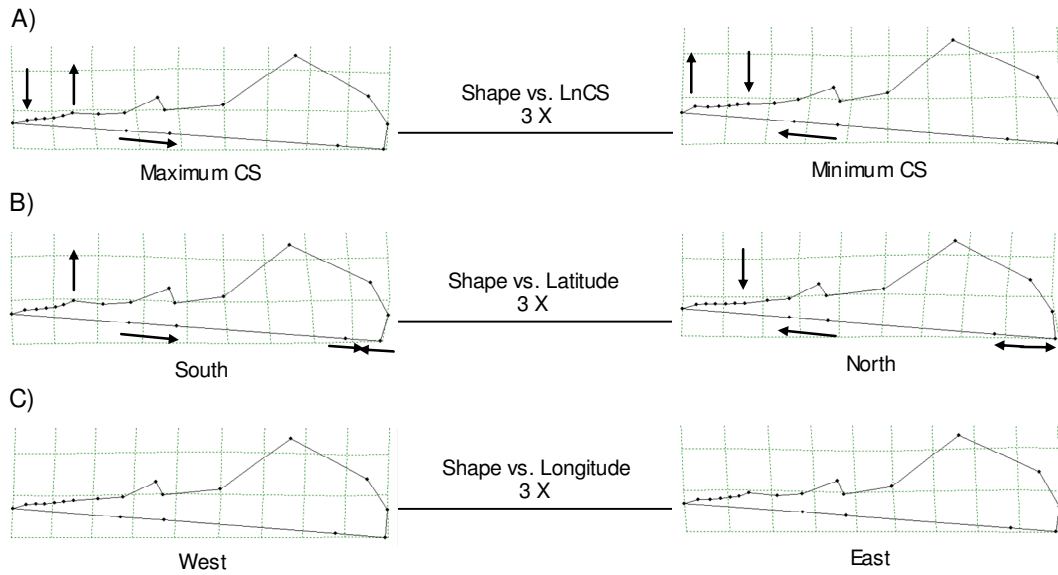
**Figure 3.11.** Multiple regressions of shape on A) Centroid Size (CS), B) latitude and C) longitude for *Sorex minutus* skulls. Transformation grids were exaggerated 3 times for purposes of visualisation. Arrows help to evaluate the relative movements of landmarks.

Latitude had a highly significant effect on skull shape and it explained a higher percentage of shape variation than CS (Latitude Goodall's $F_{34, 12750}$ = 52.6744, $p < 0.0001$, 12.31% explained). Individuals from southern latitudes had skulls of a similar shape but bigger than those from the north, with an outward movement of landmark 7 and backward movements of landmarks 18 and 19 resulting in a wider separation of upper premolars and longer maxilla and pre-maxilla, respectively, plus opposite movements of landmarks 16 and 17 resulting in a smaller foramen magnum compared to skulls from northern latitudes (Fig. 3.11B). Longitude, in contrast, had weaker effects on skull shape variation than latitude, but had stronger effects than size (Longitude Goodall's $F_{32, 18112}$ = 18.8495, $p < 0.0001$) and explained 3.8 % of shape variation (Fig. 3.11C).

The RW Analysis of skull data, as in the mandible data, showed that shape variation was not subdivided in any systematic way and there was a nearly uniform ordination of the samples around the average configuration (Fig. 3.12). However,

the first three RWs explained cumulatively 60.82% of the total variance (31.76%, 18.67% and 10.38%, respectively), about 12% more than the first three RWs of mandible data. There was little shape variation from the average configuration, but the apparent landmark movements along RW1 seem to follow the latitudinal pattern described previously. It is interesting that the first two RWs of skull data explained more variation than the first two RWs of mandible data, and that in both cases RW1 reflected the latitudinal pattern of shape variation.

The parametric and non-parametric MANOVAs revealed significant differences among groups (Wilks' $\lambda = 0.0001$, $F_{340, 3306} = 14.24$, $p < 0.0001$; non-parametric MANOVA $F_{340, 3306} = 12.05$, $p < 0.0001$), and the post-hoc tests showed that 60% and 78%, respectively, of all pairwise comparisons were significant (Appendix 8). Again, most of the non-significant differences involved comparisons with the South Italian and Iberian groups, likely a reflection of the small sample size. Including the CS as a covariate of shape variables or including *S. volnuchini* in the MANOVA made no differences to the outcome.

The discriminant function analysis of skull variation correctly classified 79.2% of the individuals to their predefined group, but using a cross-validation method only 63.1% were properly classified. With the pre-validation model, using samples with 'known' group membership, the discriminant function analysis correctly classified 90.3% of selected individuals, and with the cross-validation method it decreased to 77.3%. The first two canonical variates showed that the continental groups mostly overlapped with each other and that the island samples were more easily discriminated from the rest, especially Westray, suggesting that there were shape differences between island and continental samples (Fig. 3.13). This result was very similar to that obtained for the mandible data, where island samples could also be differentiated from continental ones.
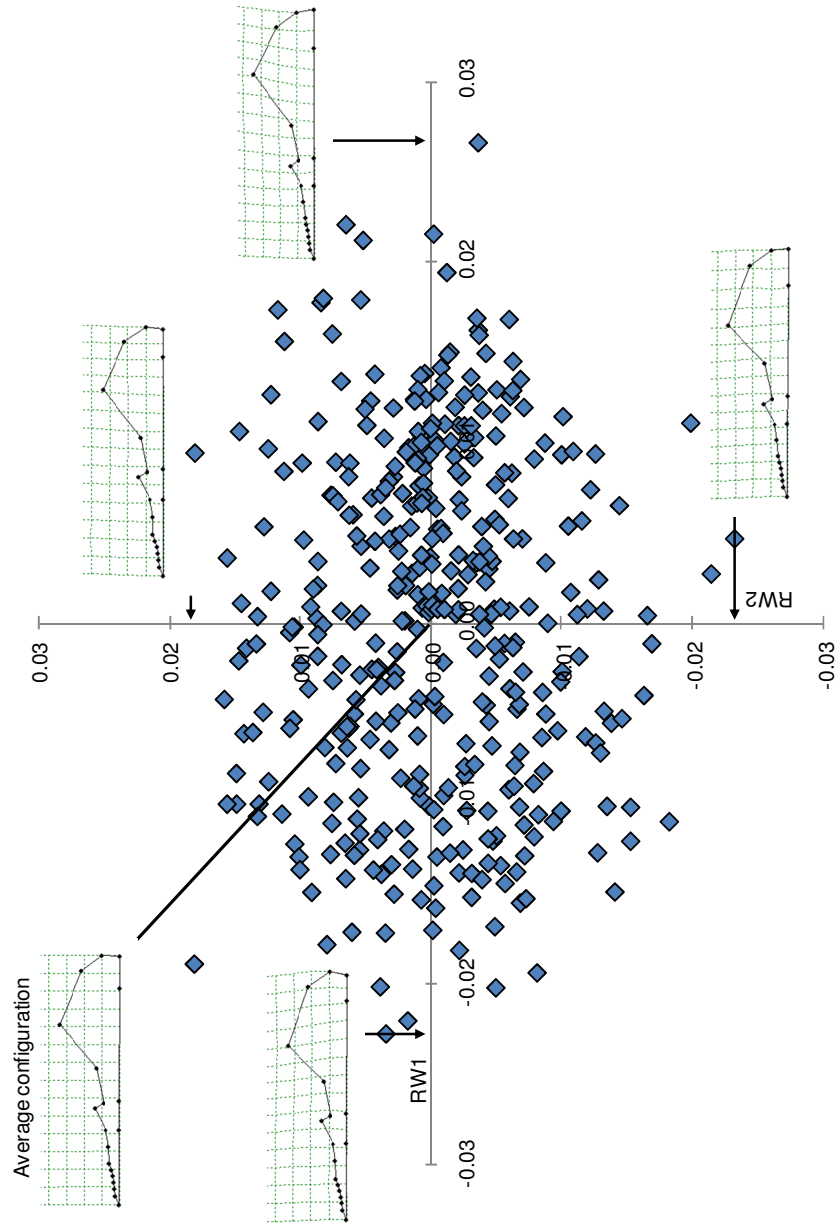
**Figure 3.12.** Relative Warp Analysis of shape variables of *Sorex minutus* skulls. The maximum and minimum transformation grids on the first two Relative Warps (RW1 and RW2), and the average configuration are shown for purposes of visualisation.

Average configuration

RW1

RW2

0.03

0.02

0.01

-0.01

-0.02

-0.03

0.03

0.02

0.01

0.00

-0.01

-0.02

-0.03
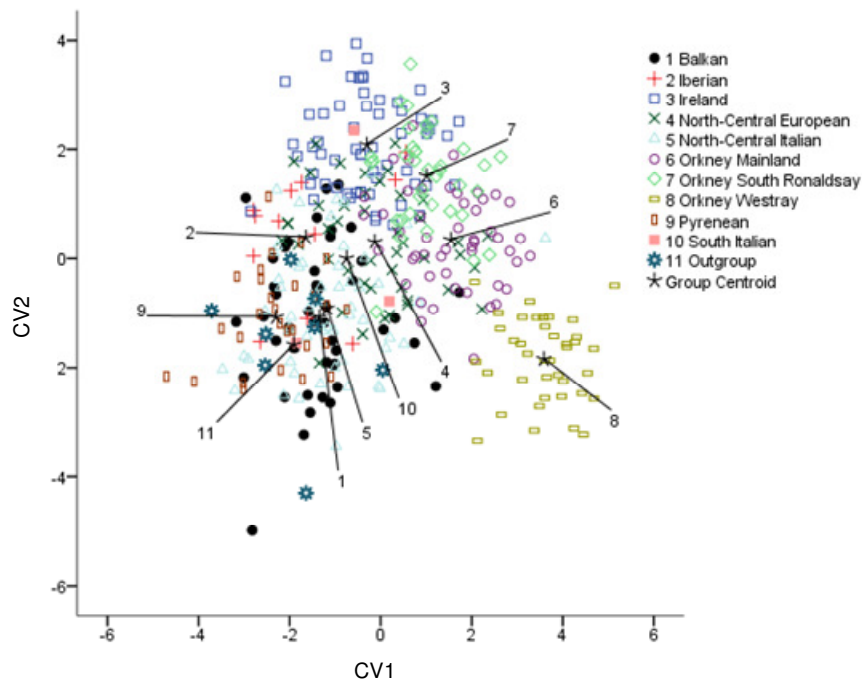
**Figure 3.13.** Canonical Variate Analysis of shape variables of *Sorex minutus* skulls showing group differences. The first two canonical variates (CV) were used to describe variation among groups.

The unrooted NJ tree of Procrustes distances among groups showed that the island groups clustered separately, with Orkney Mainland more close to other groups than to Westray and South Ronaldsay (Fig. 3.14A). Westray was again the group with the largest pairwise Procrustes distances, as with the mandible dataset. The rooted tree showed a very different topology, with a well supported cluster including all island groups, plus the North-Central European group (Fig. 3.14B). In unrooted and rooted tress, the Iberian and Pyrenean, and the Balkan and North-Central Italian groups formed two separate clusters. As with the mandible data, there was no similarity between the Procrustes and genetic NJ trees, and the amount of morphological variation did not seem to parallel the observed genetic distances (Fig. 3.14A, B).

A)

Cyt b
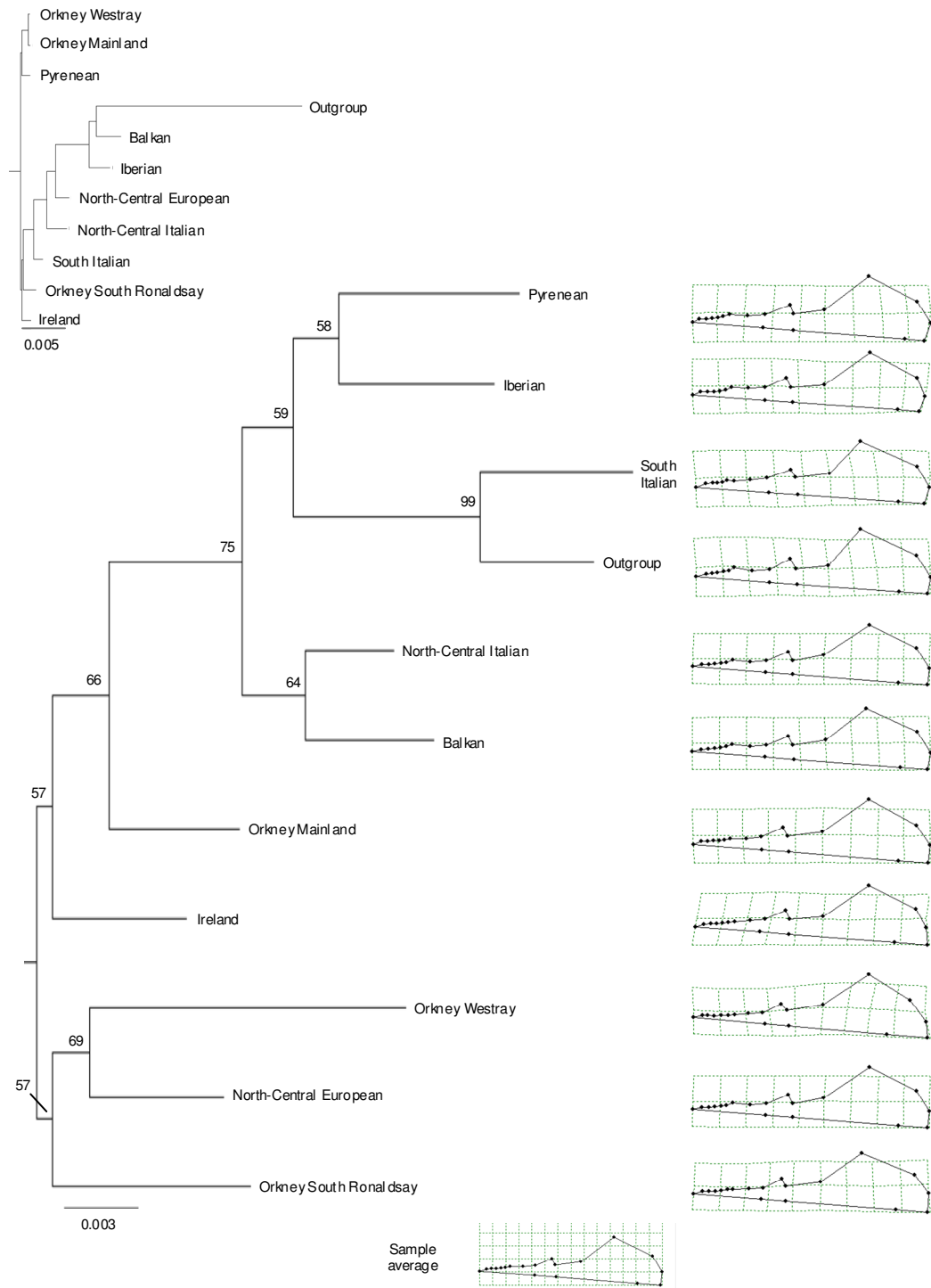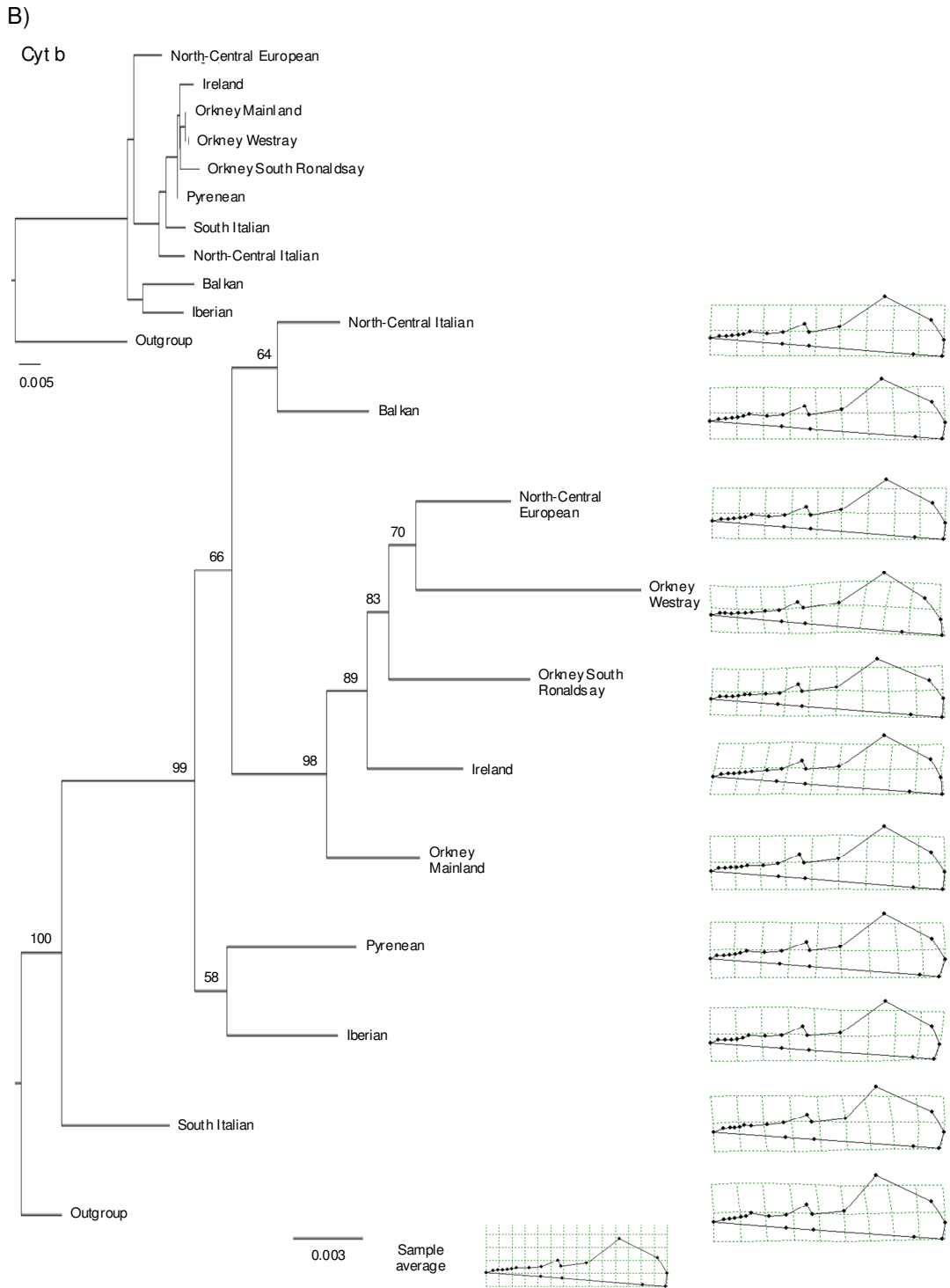


0.005



0.003

Sample
average

(Continued)

102

**Figure 3.14.** A) Unrooted and B) rooted Neighbor-Joining (NJ) trees constructed using skull Procrustes distances among *Sorex minutus* groups. Branch support is shown as bootstrap values (10,000 replications). The grids on the right correspond to shape deformation per group from the average configuration. Inset: NJ trees based on genetic distances obtained in Chapter 2, shown for purposes of comparison.

The Mantel tests for skull data revealed that there was a significant correlation between Procrustes distances and geographic distances ($R = 0.5099$, $p = 0.0021$), stronger than that found with the mandible data, and there were no correlations with genetic differentiation ($F_{ST}$; $R = -0.0955$, $p = 0.6916$) or genetic divergence (Da; $R = 0.0604$, $p = 0.4159$) among groups.

## 3.4 Discussion and conclusions.

3.4.1 Mandible and skull size variation in *Sorex minutus*

*Sorex minutus* CS had significant decreasing trends with increasing latitude, more evident for mandibles than for skulls. Southern peninsular samples, including South Italian, Iberian and Balkan groups, and the outgroup, had the biggest mandible size.

This result was contrary to that expected under Bergmann's rule, an ecogeographical rule that states that races from cooler climates tend to be larger in warm-blooded vertebrate species than races of the same species living in warmer climates (Mayr 1963). Other studies have shown a similar pattern to the one described here. For example, the condylobasal length of the skull of *S. araneus*, *S. caecutiens*, *S. tundrensis* and *S. minutus* is negatively correlated with latitude, and the strongest trend with latitude and temperature is found in *S. minutus* (Ochocińska and Taylor 2003). Ashton et al. (2000) showed that several small mammal species have significant and negative size correlations with latitude, including some species of voles (*Microtus*), kangaroo rats (*Dipodomys*) and bats (*Scotorepens*), but they only analysed a single species of shrew, *Blarina brevicauda*, which also had a negative but non-significant correlation of size with latitude. Similarly, three mainland populations of *S. trowbridgii* have decreasing cranial and mandibular dimensions with increasing latitude (Carraway and Verts 2005). Therefore, a smaller

mandible size at higher latitudes in *S. minutus* is a common trend shared with other species.

Why are shrews smaller in northern latitudes than those at southern latitudes? Lower food availability in northern/colder habitats could be the selection factor acting on small mammals, combined with lower absolute food requirements and inferior competitive ability in less productive habitats (Ochocińska and Taylor 2003). At least for *Dipodomys* and *Microtus*, seasonality and food availability appear to be the factors affecting body size (Ashton et al. 2000). Populations of *S. araneus* in Finland are up to 13% smaller on inland locations than coastal locations, where the main differences are lower winter temperatures and less snow cover at inland sites, factors associated with lower habitat productivity, which could favour selectively smaller shrews (Frafjord 2008). Moreover, bigger mandibles could reflect stronger bite force or higher mechanical potential for mastication, which could be an adaptation to more arid conditions and prey with harder exoskeletons (Strait 1993, Carraway and Verts 1994, Monteiro et al. 2003, Vega et al. 2010a). Size and shape variations of the mandible can affect the biomechanics of mastication by modifying the sites of attachment of mandible muscles (Monteiro et al. 2003).

3.4.2 Mandible and skull shape variation in *Sorex minutus*

Mandible and skull samples of *S. minutus* were not clearly differentiated by shape variation across Europe. The RW Analyses, represented graphically with the first two RWs, showed that there was no obvious differentiation of the mandible or skull samples into discrete groups. Although the first three RWs for mandibles and skulls explained 48% and 61% of the total shape variation, respectively, there was still a large amount of unexplained variation which could account for the lack of differentiation into more discrete groups of samples. The mandible and skull shape deformations along the first RW (Fig. 3.6 and Fig. 3.12) were similar to the

transformation grids obtained for the regression of shape variables on latitude and CS (Fig. 3.5 and Fig. 3.11); therefore, RW1 represented the latitudinal and allometric effects, where negative values on RW1 were equivalent to southern latitudes and larger mandible and skull size.

Dividing the sample into groups allowed me to explore the morphological differentiation under a phylogeographic scenario. MANOVAs on shape variables resulted in significantly differentiated groups. However, the CVA and classification results on mandible and skull data showed that the groups could not be differentiated or classified with confidence based on shape variables. Skulls were classified 10% better with shape variables than mandibles. This indicates that mandibles are more affected by ecophysiological factors than the ventral side of skulls (Monteiro et al. 2003, Caumul and Polly 2005). In this study, no morphological structure recovered the phylogenetic relationships detected with the cyt *b* analysis, and the morphological variation among the groups increased with geographic distance, not with genetic distance or differentiation for mandible and skull data. As expected from intra-specific studies such as this, the molecular differentiation was more pronounced than the morphological differentiation. Contrary to this, Carraway and Verts (2005) found, with a traditional morphometric analysis using linear measurements, that cranial and mandible characters recover relatedness within *S. trowbridgii*, and that mandibles alone classify specimens correctly to their groups selected a priori; although they also distinguish mainland from insular taxa.

Interestingly, island samples of *S. minutus* grouped somewhat separately from the rest in the CVA. Moreover, island samples tended to cluster together in NJ trees based on Procrustes distances. Latitudinal and allometric effects could explain this pattern: all the island samples were obtained from higher latitudes than continental ones, and specifically Westray samples, with the highest latitude (59° north latitude), showed the strongest shape differentiation. A more localized study of

the genetic and morphological differentiation of island and continental samples is presented in Chapter 4.

Size, not shape, was the principal and most obvious difference in mandibles and skulls among the European regions. In general, there appeared to be isotropic shape variation for mandibles and skulls, most evident in landmarks 2 to 15 of mandibles and landmarks 2 to 12, 14 and 15 of skulls (e.g., see the distribution of landmark coordinates in Fig. 3.2). Isotropic shape variation occurs when the variation at each landmark is the same in all directions and independent among landmarks, meaning that the scatter of landmark positions around the average is circular and has the same spread at each landmark (Klingenberg and Monteiro 2005). However, the shape variation among regions is of interest to compare with patterns of genetic variation.

The evidence provided here supports the conclusion that latitude and island evolution are the driving factors of size and shape variation in *S. minutus* mandibles and skulls, probably reflecting selection forces or size and shape adaptations to different habitats, food availability and environmental conditions (Fadda and Corti 2001). Overall, both datasets showed similar patterns of shape variation, but latitudinal and size effects were more evident with mandibles than skulls. This is not unexpected because mandibles usually show more ecophenotypic variation than the ventral side of the skull. A detailed analysis of size and shape variation in an island setting is presented in Chapter 4.

3.4.3 Taxonomic implications

In the previous chapter (Chapter 2), I described four phylogroups based on cyt *b*, including the Balkan, Iberian, South Italian and North-Central European, which roughly matched the geographic distribution of four subspecies of *S. minutus* described by Hutterer (1990): *S. m. gymnurus* (Greece), *S. m. carpetanus* (Iberia),

*S. m. lucanius* (Southern Italy) and *S. m. minutus* (Northern-Central Europe and Siberia), respectively.

The geometric morphometric analyses presented here did not parallel the phylogeographic groupings, but there was important size and shape variation. Populations in southern peninsulas had significantly bigger mandibles than those from Northern-Central Europe, and there were also significantly different groups despite the small amount of shape variation. Similarly, the three subspecies from the southern peninsulas described by Hutterer (1990) are bigger than *S. m. minutus* from Northern-Central Europe. Therefore, this study provided morphological and genetic evidence that support the subspecific status for a Balkan group (equivalent to *S. m. gymnurus*), an Iberian group (equivalent to *S. m. carpetanus*) and a South Italian group (equivalent to *S. m. lucanius*), and a morphologically smaller and genetically distinct North-Central European-Siberian group (equivalent to *S. m. minutus*), all of which could have evolved allopatrically during the glacial/interglacial periods.

The subspecific status of the North-Central Italian and Pyrenean groups is uncertain, however. These two groups were closely related in the cyt *b* analysis and diverged 0.225 MYA according to the molecular clock applied in section 2.3.4, and had a similar average mandible and skull size, but were significantly bigger than the North-Central European group. It is expected that recently separated species/subspecies may be genetically different but not well-differentiated morphologically. This is evident from the observed size and shape similarities between *S. volnuchini* and *S. minutus* samples, which are genetically very distinct. It is possible that the North-Central Italian and Pyrenean groups correspond to two distinct subspecies, one equivalent to *S. m. becki* (Alps) and the other one not yet described.

# Chapter 4

# Island phylogeography and morphometrics of *Sorex minutus*

### 4.1 Introduction

During the Last Glacial Maximum (LGM; 23 – 18 KYA), the coldest stage of the Devensian (Weichselian) glaciation, ice sheets covered most of Britain and Ireland, while only the southern-most regions were unglaciated, but presented tundra and permafrost conditions unsuitable for most of the present mammal fauna (Yalden 1982, 2007). The timing of the migration of small mammals from continental Europe into Britain must have coincided then with the presence of an exposed landbridge between Northern France and Southern Britain; however, the natural colonisation of Ireland and other outer islands of the British Isles from source populations in mainland Britain might have not been possible because of the absence of a landbridge interconnecting them (Yalden 1982, Stuart 1995, Lambeck and Purcell 2001). Thus, the presence of the Eurasian pygmy shrew *Sorex minutus* on Britain, Ireland and other outer islands of the British Isles makes it an interesting model for understanding natural and human-mediated island colonisation and evolution.

Previous work on *S. minutus* based on microsatellite data provided evidence for a population expansion and colonisation event in Ireland ca. 6 KYA (7 – 4 KYA), in the Neolithic (McDevitt et al. 2009). The results provided in Chapter 2 indicate a source area for the Irish pygmy shrew in Northern Spain, given the haplotype similarity between Navarra, Spain and those from Ireland, all of which belong to a widespread Pyrenean lineage distributed in Northern Spain, Western and Southern France. Searle et al. (2009) showed that there are individuals of *S. minutus* in

mainland Britain and islands off the coast of Britain, which also belong to the Pyrenean lineage, and represent a 'Celtic fringe' only found in the western and northern regions of mainland Britain and in some offshore islands. Archaeological evidence indicates that Neolithic people travelled by boat between mainland Britain and nearby islands, and also among the Orkney islands (Phillips 2004), and the same applies to more recent human occupants. Therefore; it is expected that small mammals, including *S. minutus*, were sometimes accidentally carried from one region to another by boat, probably hiding in livestock fodder.

It is not only colonisation history of islands that has impacted the genetic and morphological diversity seen on them; island size, selective pressures and geographical isolation are also important (Lomolino 2005). Several studies indicate that mammal populations on islands usually show lower population size, inbreeding and lower levels of genetic diversity than continental populations (Frankham 1998), and morphological traits evolve more rapidly compared to continental counterparts, from a few decades to several thousand years (Millien 2006). For example, White and Searle (2007, 2008a) showed that island populations of *S. araneus* have lower microsatellite expected heterozygosity and lower mean cytochrome (cyt) *b* nucleotide diversity than mainland populations, and that there is morphological differentiation of island populations with variation of mandibles between islands strongly predicted by climate (White and Searle 2009). A meta-analysis of rodents indicates that microevolution and selection cause rapid substantial shifts in morphology among insular rodents shortly after island colonisation (Pergams and Ashley 2001), although other have argued the importance of founder events (Berry 1996). *Marmota vancouverensis*, the only insular species of the genus *Marmota*, has low genetic divergence but strong morphological differences from continental species that evolved rapidly following its isolation after the LGM (Nagorsen and Cardini 2009).

Here, I analyse the genetic and morphological diversity of pygmy shrews within Ireland, the Orkney islands and the periphery of mainland Britain, all belonging to a Pyrenean mitochondrial lineage described previously, and compare it with a continental European population belonging to the same lineage. In this way, island colonisation history and evolution can be examined within a defined genetic lineage. In particular, the objectives are: 1) to investigate the genetic diversity, origin and migration rates among the continental Pyrenean and the Orkney, Ireland and British Celtic fringe populations from the viewpoint of island colonisation and evolution, 2) to evaluate the morphological variation among and within the islands in comparison to the continental population of a phylogenetically related sample, and 3) to evaluate the specific influence of island size on the morphology of *S. minutus*.

## 4.2 Materials and methods

4.2.1 Samples for genetic and morphological analysis

For the genetic analysis, I used the samples and available sequences that clustered in the full cyt *b* Pyrenean lineage described in Chapter 2. This geographically widespread lineage included samples from the Cantabrian mountain range in Northern Spain, the Pyrenees in North-Eastern Spain and Southern France, Western and Central France, and from the island of Belle Île en Mer (off the coast of Western France), plus all samples that formed the 'Ireland', 'Orkney islands' and British 'Celtic fringe' groups.

I divided the full cyt *b* Pyrenean lineage into four main groups (Fig. 4.1): the 'Continental Pyrenean' (for simplicity now named as 'Continental', but also including the few samples from the island of Belle Île), 'Ireland', 'Orkney islands' and the British 'Celtic fringe'. The samples from Orkney belong to the Celtic fringe described by Searle et al. (2009), but for the purpose of this study they were considered as a separate group.
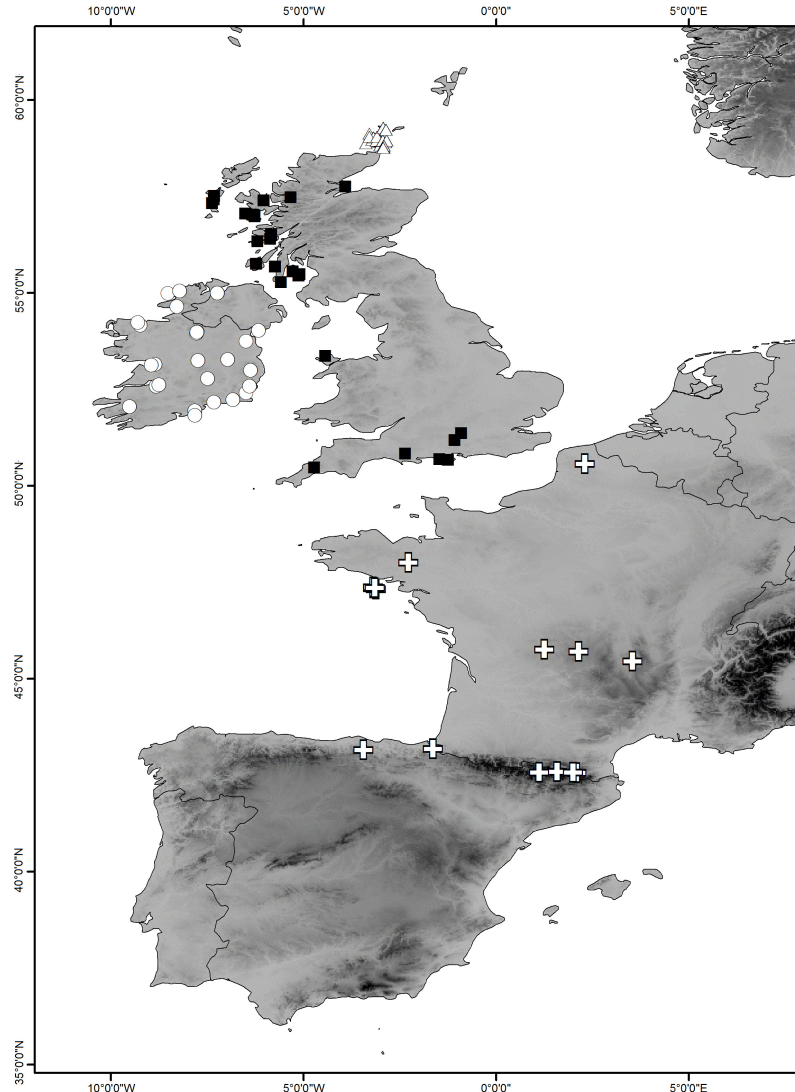
**Figure 4.1.** Samples of *Sorex minutus* for genetic analysis. ✛ Continental, ■ British Celtic fringe, ◯ Ireland and △ Orkney islands.

The genetic sample from Ireland was subdivided into nine regions: Cork, Derry, Gweedore, Limerick, Wexford, Galway, Kerry, Cloghan and Kildare. The Orkney sample was subdivided into: Orkney Mainland, Westray, South Ronaldsay and Hoy, and into regions or localities within these islands, including three regions within Orkney Mainland (named Mainland 1, Mainland 2 and Mainland 3), two

localities within Westray (Loch of Swartmill and Ness), two localities within South Ronaldsay (Grimness and Windwick), and only one locality for Hoy.

For the geometric morphometrics analysis, I included the mandible and skull samples that formed the 'Pyrenean', 'Ireland', 'Orkney Mainland', 'Westray' and 'South Ronaldsay' morphological groups described in Chapter 3 (see also Fig. 3.1). The Pyrenean sample was divided into 'Navarra' (Northern Spain), 'continental France' (mostly from Central France) and 'Belle Île' (from the island of Belle Île en Mer, off the coast of Western France). Ireland was subdivided into three regions: Cloghan, Derry and Gweedore. The Orkney islands were further subdivided into regions or localities: Orkney Mainland was divided into Mainland 1 and Mainland 2, Westray was divided into Ness and Loch of Swartmill, and South Ronaldsay was divided into Grimness and Windwick.

To investigate island size as a factor affecting the morphology of *S. minutus*, I explored size and shape variation dividing the mandible and skull datasets into four different categories: 1) 'Continental' (grouping samples from Northern Spain and France), 2) 'Large' size islands (Ireland), 3) 'Medium' size islands (Orkney Mainland), and 4) 'Small' size islands (grouping samples from South Ronaldsay, Westray and Belle Île).

4.2.2 Sequence diversity, divergence and differentiation

Sequence diversity, divergence and genetic differentiation parameters were calculated in DnaSP version 5.10 (Librado and Rozas 2009), as explained in Chapter 2. I calculated the genetic ($\pi$) and haplotype (Hd) diversity values for the Continental group, and for the islands and regions within islands. Then, I obtained the genetic divergence (Da) among the main groups (Continental, Orkney Mainland, Westray, South Ronaldsay, Hoy and Ireland), and among the regions within the main groups.

The pairwise geographic distances (in Km) were calculated among the main continental-island groups, and among the regions within islands and the Continental group. Mantel tests were performed in Arlequin version 3.1 (Excoffier et al. 2005) to test the significance of the correlation between the pairwise Da values with the pairwise geographic distances among groups and among regions within groups separately. The significance was obtained through a permutation procedure (10,000 replications) used to create an empirical null distribution.

The structure of the genetic diversity was evaluated using Arlequin with an Analysis of Molecular Variance (AMOVA). I implemented a hierarchical analysis dividing the total variance into three groups (Continental, Orkney and Ireland), and into regions within groups (Mainland 1, Mainland 2, Mainland 3, Loch of Swartmill, Ness, Grimness and Windwick for the Orkney archipelago, and Cork, Derry, Gweedore, Limerick, Wexford, Galway, Kerry and Kildare for Ireland).

Finally, I used the Median-Joining (MJ) phylogenetic network presented in Chapter 2 (Fig. 2.3C) to represent the genetic relationships among continental and island haplotypes that formed the full Pyrenean phylogroup.

4.2.3 Estimation of migration routes

The population parameter 'theta' ($\theta$ = 2Neu) and the migration rates between the Continental, Celtic fringe, Ireland and Orkney groups were calculated using Lamarc version 2.1.3 (Likelihood Analysis with Metropolis Algorithm using Random Coalescence), which implements a coalescent method (Kuhner 2006). The effective population size (Ne) was calculated solving for $\theta$, where u = 2$\mu$k , $\mu$ is the mutation rate per nucleotide (assuming a mutation rate per nucleotide of 2% per million years, Wilson et al. 1985) per generation (considering one generation per year for *S. minutus*) and k is the number of nucleotide sites (1110 bp for cyt *b* data) (Rogers 1995). In this program, the genealogical relationships among all pair of individuals

are estimated revealing coalescent-based information on the genealogies, and a random effect is incorporated using the Metropolis-Hastings Markov Chain Monte Carlo (MCMC) sampling. The method gives an approximation to the actual value being estimated, but an error range around the Most Probable Estimate (MPE) is generated, which is equivalent to a confidence interval (Kuhner 2006).

For running the program, I selected 10 initial runs of 10,000 generations sampled every 20 generations and a burn-in of 250, followed by two final runs of one million generations sampled every 100 generations with a burn-in of 2500. The results were averaged over three replicates (using the same parameters) and convergence was visualised in Tracer version 1.4.1 (Drummond and Rambaut 2007). I kept default values for the Bayesian start parameters, but set the evolutionary model to the General Time Reversible (GTR).

4.2.4 Geometric morphometric analysis

The mandible and skull samples were statistically analysed in SPSS/PASW Statistics version 17 (IBM), PAST version 1.97 (Hammer et al. 2001) and the 'tps-series' software (by FJ Rohlf, available at http://life.bio.sunysb.edu/morph/). The landmarks used in this study were the same as in Chapter 3. The landmark configurations were scaled to unit Centroid Size (CS; natural log-transformed), translated and rotated using the Generalised Procrustes Analysis (GPA). CS was used for the analysis of size variation and the weight matrix ($W$) was used for the analysis of shape variation. Descriptive statistics were obtained for the mandible and skull samples.

Mandible and skull size differences among continental and island groups were evaluated by an Analysis of Variance (ANOVA) and visualized with box plots. Levene's tests were conducted to detect heteroscedasticity, but the ANOVA is robust to this (Zelditch et al. 2004). I performed a Tukey–Kramer post-hoc test on

CS as it allows for unequal sample size (Sokal and Rohlf 1995). In the case of a significant departure from homogeneity of variances, I performed a non-parametric Kruskal–Wallis test and a Tamhane T2 non-parametric post-hoc test for the analysis of CS variation.

Mandible and skull shape differences among continental and island groups were evaluated with a Multivariate Analysis of Variance (MANOVA) on $W$, followed by Hotelling $T^2$ tests. Levene's tests were also performed for each shape variable among groups and, if significant deviations from the assumption of homogeneity of variances were found, non-parametric MANOVAs among groups were also performed.

The Procrustes distances among the average configurations of the groups were then computed with the program tpsSmall. I performed Mantel tests in PAST to test the significance of the correlation between pairwise Procrustes distances and geographic distances and Da among groups (the significance was obtained with 10,000 permutations).

A Discriminant Analysis (DA) was performed in SPSS to predict group membership using a cross-validation method and to differentiate among the groups by a Canonical Variates Analysis (CVA), as explained in Chapter 3. The first two canonical variates were used to graph the samples separated by group membership. A similar approach was used to discriminate and classify the mandible and skull samples divided into island size categories.

## 4.3 Results

4.3.1 Genetic diversity and structure

The Continental (n = 20) and Ireland groups (n = 94) had high cyt $b$ haplotype diversities (Hd = 0.9840 and Hd = 0.8920, respectively) compared to the full Orkney islands sample (n = 119, Hd = 0.7720), or to any of the individual Orkney

islands with adequate sample sizes (Table 4.1). However, nucleotide diversity was always low for island samples compared to the Continental group (Table 4.1). The largest island of the Orkney archipelago, Orkney Mainland, had the highest nucleotide and haplotype diversity values. Interestingly, South Ronaldsay had very low genetic diversity and there were only two haplotypes in a sample of 40 individuals, while Westray, the northern-most island sampled, had only one haplotype in 33 sampled individuals.

**Table 4.1.** Cytochrome *b* sequence diversity of *Sorex minutus* groups of the Pyrenean lineage.

| Groups | n | Haplotypes | Hd | π | K | S |
|---|---|---|---|---|---|---|
| Continental | 20 | 17 | 0.9840 | 0.0054 | 5.9370 | 33 |
| Orkney islands | 119 | 11 | 0.7720 | 0.0027 | 3.0140 | 17 |
| Orkney Mainland | 44 | 8 | 0.7550 | 0.0013 | 1.4790 | 9 |
| Orkney Mainland1 | 29 | 5 | 0.5120 | 0.0010 | 1.0940 | 6 |
| Orkney Mainland2 | 10 | 4 | 0.8000 | 0.0012 | 1.3330 | 4 |
| Orkney Mainland3 | 5 | 2 | 0.4000 | 0.0004 | 0.4000 | 1 |
| South Ronaldsay | 40 | 2 | 0.1420 | 0.0001 | 0.1420 | 1 |
| Grimness | 21 | 2 | 0.0950 | 0.0001 | 0.0950 | 1 |
| Windwick | 19 | 2 | 0.1990 | 0.0002 | 0.1990 | 1 |
| Westray | 33 | 1 | 0.0000 | 0.0000 | 0.0000 | 0 |
| Loch of Swartmill | 19 | 1 | 0.0000 | 0.0000 | 0.0000 | 0 |
| Ness | 14 | 1 | 0.0000 | 0.0000 | 0.0000 | 0 |
| Hoy | 2 | 2 | 1.0000 | 0.0018 | 2.0000 | 2 |
| Ireland | 94 | 42 | 0.8920 | 0.0020 | 2.2180 | 53 |
| Cork | 10 | 7 | 0.8670 | 0.0043 | 4.7560 | 15 |
| Derry | 11 | 2 | 0.3270 | 0.0003 | 0.3270 | 1 |
| Gweedore | 6 | 3 | 0.6000 | 0.0011 | 1.2000 | 3 |
| Limerick | 7 | 5 | 0.8570 | 0.0013 | 1.4290 | 5 |
| Wexford | 11 | 4 | 0.6730 | 0.0012 | 1.3090 | 4 |
| Galway | 5 | 5 | 1.0000 | 0.0025 | 2.8000 | 7 |
| Kerry | 5 | 3 | 0.7000 | 0.0007 | 0.8000 | 2 |
| Kildare | 5 | 3 | 0.7000 | 0.0013 | 1.4000 | 3 |
| Cloghan | 3 | 2 | 0.6667 | 0.0006 | 0.6667 | 1 |

n = Number of sequences, Hd = Haplotype diversity, π = Nucleotide diversity (Jukes and Cantor 1969), K = Average number of nucleotide differences, S = Number of polymorphic (segregating) sites.

There was no evidence of isolation by distance for the whole dataset ($R = 0.0204$, $p = 0.3933$), because of the weak and non-significant correlation of geographic distances and genetic divergence values among the Continental and island samples, as expected for populations evolving in an island setting. Moreover, there were very small divergence values, as expected, and no evidence of isolation by distance among localities within Ireland ($R = 0.3597$, $p = 0.0736$) or among localities of the Orkney islands ($R = 0.2105$, $p = 0.0809$).

The genetic diversity of the whole dataset was significantly structured ($p = 0.0001$), with most of the variation found among the Continental and island groups (43%), followed by the variation among localities within the groups (34%) and within localities (22%). For the Orkney archipelago, there was a strong and significant genetic structure ($p < 0.0001$), with more variation found among islands (87%) than within (13%). In contrast, there was less variation among localities (26%) than within localities (74%) in Ireland, but there was also a significant structure ($p < 0.0001$). The genetic structure for the Continental group was not calculated because of low sample size per locality. These results showed that populations of *S. minutus* within Ireland and the Orkney islands are subdivided, especially so on the Orkney islands, although there were small values of genetic diversity there.

The MJ network showed that the haplotype ESBa9709 (from near Barcelona, Spain), was the most basal of the Pyrenean haplotypes to the full network and connected the Pyrenean lineage to the other Eurasian phylogroups. From this haplotype, the other more derived Continental, Orcadian, Irish and Celtic fringe haplotypes extended with one or more mutational steps (Fig. 4.2).
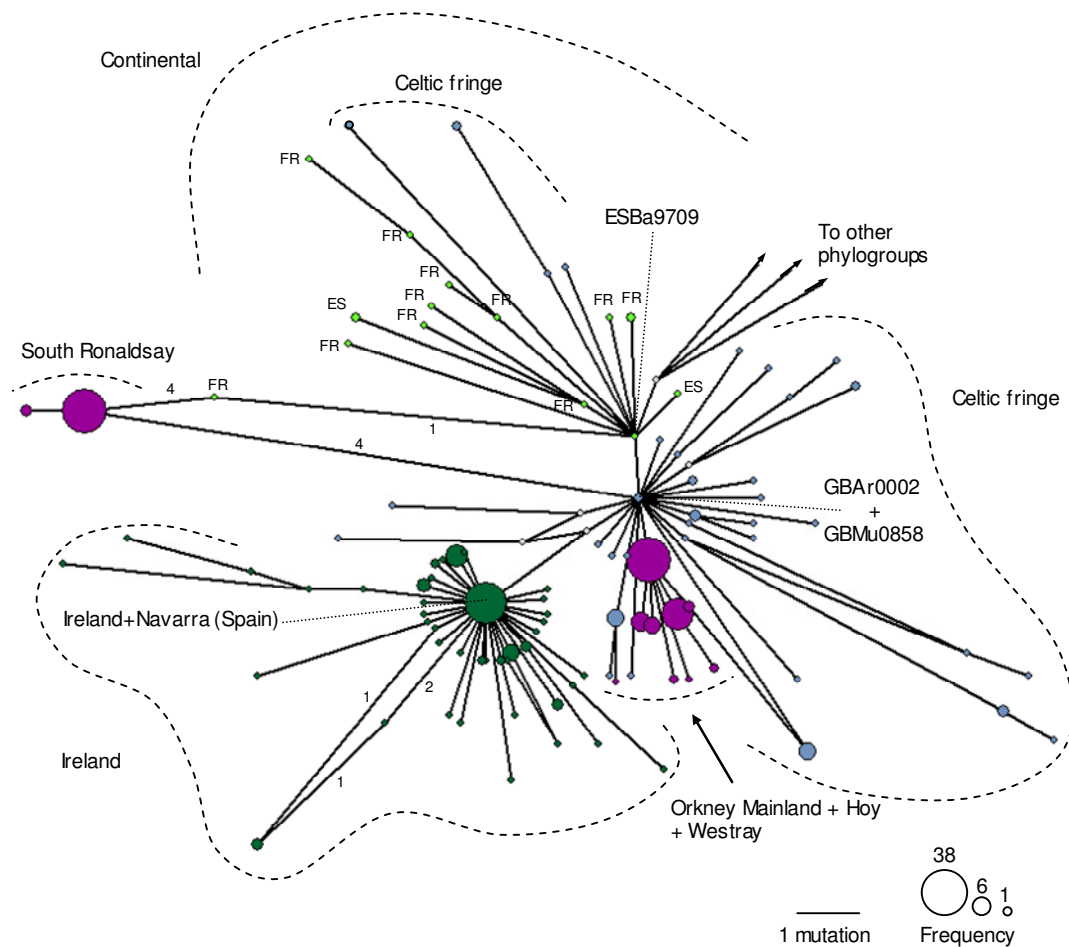
**Figure 4.2.** Median-Joining network of cytochrome *b* data showing the relationships among *Sorex minutus* haplotypes belonging to the Pyrenean lineage. The Pyrenean lineage is divided into a 'Continental' group (light green), a 'Celtic fringe' group distributed on Britain and some islands off the coast of Britain (light blue), excluding the Orkney islands, 'Ireland' (dark green) and 'Orkney islands' groups (purple). Pointed lines indicate the central (ancestral) haplotypes from where all other haplotypes derive from (see text for explanation). Some loops in the network are not in scale and mutations are shown with numbers. For the 'Continental' group, the country of origin for the haplotypes is shown (FR = France, ES = Spain).

All haplotypes from the Orkney islands were most closely connected to a central haplotype from the Celtic fringe (GBAr0002 and GBMu0858); therefore, *S. minutus* from the Orkney islands can be considered part of the Celtic fringe. Orkney Mainland, the largest island of the Orkney archipelago, was represented by eight haplotypes (n = 44). There were only two haplotypes in South Ronaldsay (n = 40)

separated by four mutations, and Westray had only one haplotype (n = 33), identical to a highly frequent haplotype from Orkney Mainland. All haplotypes from Ireland derived from one Irish central haplotype, which included two Spanish samples (from Navarra, Northern Spain) and 29 Irish samples (from several localities). All the other haplotypes from Navarra, Spain were directly connected with the haplotype ESBa9709. The Irish and Celtic fringe central haplotypes were connected to each other by three mutational steps and one hypothetical haplotype. No Irish haplotype connected directly to the Celtic fringe or to other Continental haplotypes; however, three Celtic fringe haplotypes were directly connected to ESBa9709, showing a close relationship between Continental and Celtic fringe haplotypes. There were several divergent Irish haplotypes without any obvious geographical distribution.

4.3.2 Migration routes

The MPEs for effective population size, θ values and migration rates among the Continental, Celtic fringe, Ireland and Orkney islands groups showed convergence of the three replicates after several runs, but the 95% confidence intervals were wide so these results have to be taken with caution (Table 4.2). The cyt *b* sequences did not contain enough information to provide adequate estimates; therefore, the MPEs should be interpreted as general tendencies or trends, not as precise values.

The Continental group had the highest θ values and effective population size, as expected from a genetically diverse continental source, followed by the Celtic fringe group. The Celtic fringe group was composed of samples from different localities from Britain and offshore islands, so the high θ and effective population size values found there likely resulted because of pooling genetically distinct populations into one single group; however, this grouping likely reflects the genetic diversity of what was once a more widespread lineage in mainland Britain. Ireland

had intermediate θ and effective population size values, compared to the Continental and Orkney islands groups, as expected for a population on a large island and genetically differentiated into subpopulations. The Orkney islands group had the smallest θ and effective population size values. Considering that *S. minutus* is abundant on the Orkney islands, these estimates resulted from the low diversity of the molecular marker used, likely a reflection of a relatively recent colonisation event by a small number of founders.

**Table 4.2.** Demographic parameters obtained for *Sorex minutus* Pyrenean groups based on cytochrome *b* data.

|   |   | Continental | Celtic fringe | Ireland | Orkney |
|---|---|---|---|---|---|
| θ | MPE | 0.0137 | 0.0105 | 0.0065 | 0.0014 |
|   | Lower | 0.0075 | 0.0057 | 0.0032 | 0.0004 |
|   | Upper | 0.0262 | 0.0217 | 0.0128 | 0.0032 |
| Ne | MPE | 154.0428 | 117.7928 | 73.1531 | 15.7545 |
|   | Lower | 84.5158 | 64.2004 | 35.8559 | 5.0338 |
|   | Upper | 295.6306 | 244.6622 | 144.0090 | 36.4865 |

MPE = Most Probable Estimate, Lower corresponds to lower 95% CI, Upper corresponds to upper 95% CI, θ= 2Neu (u = 2μk).

The migration rates showed a trend from continental to island regions and from larger islands to smaller islands (Fig 4.3). Continental Europe appeared as a source area for the Ireland, Celtic fringe and Orkney islands groups because of the higher migration rates from the continent to islands. However, the migration rates between the Celtic fringe and the Continental groups were almost equal, slightly favouring the Continental group as a source area. The migration rates between the Celtic fringe and Ireland groups were almost equal, showing a slight tendency of the former as a source area for the later. The Celtic fringe had the strongest migration rate onto the Orkney islands, which could be explained by the geographical proximity, and represent the most likely source for the Orkney pygmy shrew.
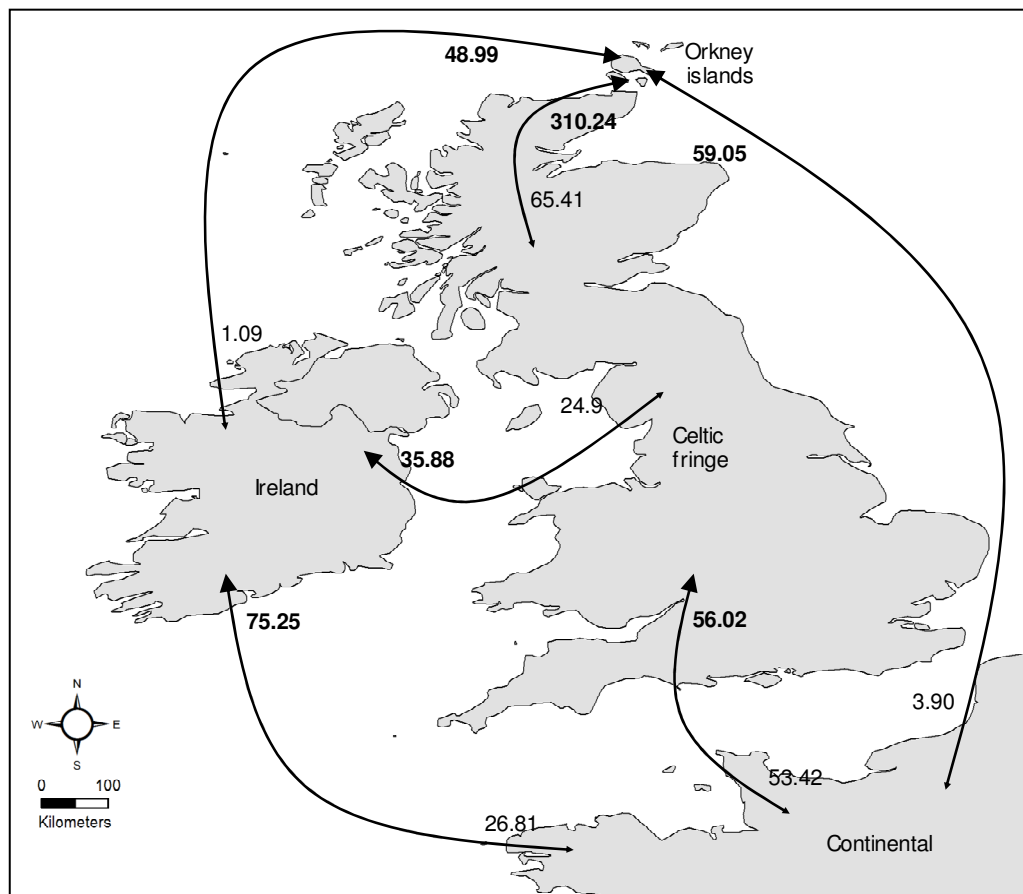
**Figure 4.3.** Migration rates and routes of *Sorex minutus* among the Continental, Celtic fringe, Ireland and Orkney islands groups estimated from cytochrome *b* data. Values next to arrows are the estimated migration rates from one group to the other. Larger arrow heads and bold numbers represent the migration rate trends, mainly from continent to island and from larger island to smaller island.

### 4.3.3 Geometric morphometrics findings

The decreasing trend of CS with increasing latitude, as detected for the full Eurasian sample (Chapter 3), was not reflected convincingly in the reduced dataset examined here, although two of the more southerly samples included here (Navarra and Belle Île) did have the highest values (Table 4.3, Fig. 4.4A). Levene's test was not significant for the mandible dataset (Levene's statistic$_{11, 260}$ = 0.8090, $p$ = 0.6310), and the ANOVA ($F_{11, 260}$ = 27.2100, $p < 0.0001$) and parametric post-hoc tests showed that 47% of the pairwise comparisons of mandible CS were significant

(Appendix 9). Among the populations from the Orkney islands, Westray Ness was the only one that was not significantly different from any other island group.

The CS of skull groups showed no apparent trend with latitude (Table 4.4, Fig. 4.4B), as expected from the full European sample analysed in Chapter 3. Mainland Orkney had the largest skull size variation and the two groups from Westray had the biggest skull CS. The test of homogeneity of variances for skull CS was significant (Levene's statistic$_{9, 204}$ = 7.4460, $p$ < 0.0001), so a non-parametric ANOVA was performed and it was significant (H$_9$ = 28.7680, $p$ < 0.0001). The only significant post-hoc comparisons were among Westray Ness, Cloghan and Derry (Appendix 10).

**Table 4.3.** Centroid Size statistics for *Sorex minutus* mandibles of the Pyrenean lineage.

| Group | n | mean | median | Lower | Upper | SD | SE |
|---|---|---|---|---|---|---|---|
| Navarra | 28 | 2.4058 | 2.4052 | 2.3984 | 2.4132 | 0.0192 | 0.0036 |
| Continental France | 30 | 2.3361 | 2.3406 | 2.3292 | 2.3430 | 0.0192 | 0.0036 |
| Belle Île | 19 | 2.3784 | 2.3745 | 2.3715 | 2.3853 | 0.0143 | 0.0033 |
| Ireland Cloghan | 20 | 2.3610 | 2.3654 | 2.3497 | 2.3722 | 0.0241 | 0.0054 |
| Ireland Derry | 23 | 2.3459 | 2.3504 | 2.3380 | 2.3537 | 0.0182 | 0.0038 |
| Ireland Gweedore | 20 | 2.3393 | 2.3364 | 2.3285 | 2.3501 | 0.0231 | 0.0052 |
| O SR Windwick | 20 | 2.3415 | 2.3442 | 2.3353 | 2.3478 | 0.0133 | 0.0030 |
| O SR Grimness | 20 | 2.3366 | 2.3368 | 2.3297 | 2.3434 | 0.0147 | 0.0033 |
| O Mainland 2 | 39 | 2.3538 | 2.3564 | 2.3468 | 2.3608 | 0.0215 | 0.0034 |
| O Mainland 1 | 13 | 2.3687 | 2.3664 | 2.3537 | 2.3837 | 0.0248 | 0.0069 |
| O W Ness | 20 | 2.3551 | 2.3533 | 2.3479 | 2.3623 | 0.0154 | 0.0034 |
| O W Swartmill | 20 | 2.3629 | 2.3590 | 2.3521 | 2.3737 | 0.0231 | 0.0052 |

O = Orkney, SR = South Ronaldsay, W = Westray, Lower = Lower 95% CI, Upper = Upper 95% CI, SD = Standard deviation, SE = Standard error.
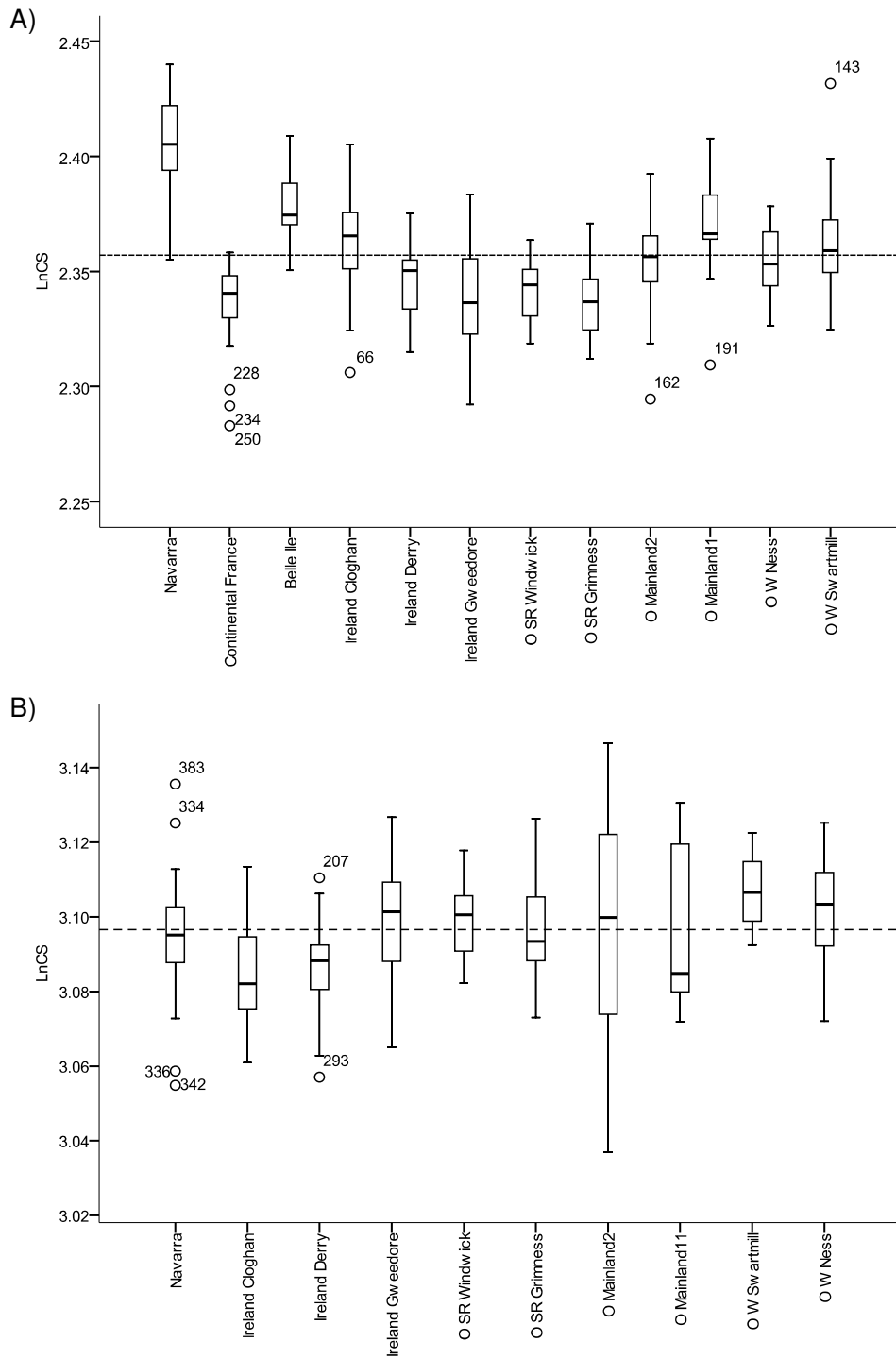
**Figure 4.4.** Box plots of A) mandible and B) skull Centroid Size (CS; natural log-transformed) of *Sorex minutus* samples ordered latitudinally (with south to the left).

**Table 4.4.** Centroid Size statistics for *Sorex minutus* skulls of the Pyrenean lineage.

| Group | n | mean | median | Lower | Upper | SD | SE |
|---|---|---|---|---|---|---|---|
| Navarra | 28 | 3.0947 | 3.0951 | 3.0879 | 3.1014 | 0.0174 | 0.0033 |
| Ireland Cloghan | 20 | 3.0847 | 3.0821 | 3.0777 | 3.0917 | 0.0150 | 0.0034 |
| Ireland Derry | 21 | 3.0865 | 3.0883 | 3.0805 | 3.0925 | 0.0132 | 0.0029 |
| Ireland Gweedore | 19 | 3.0998 | 3.1014 | 3.0920 | 3.1075 | 0.0161 | 0.0037 |
| O SR Windwick | 18 | 3.0994 | 3.1006 | 3.0941 | 3.1048 | 0.0108 | 0.0025 |
| O SR Grimness | 19 | 3.0972 | 3.0934 | 3.0902 | 3.1041 | 0.0144 | 0.0033 |
| O Mainland 2 | 37 | 3.0989 | 3.0999 | 3.0889 | 3.1090 | 0.0300 | 0.0049 |
| O Mainland 1 | 13 | 3.0966 | 3.0849 | 3.0833 | 3.1100 | 0.0222 | 0.0061 |
| O W Swartmill | 20 | 3.1070 | 3.1065 | 3.1025 | 3.1115 | 0.0095 | 0.0021 |
| O W Ness | 19 | 3.1016 | 3.1034 | 3.0939 | 3.1092 | 0.0158 | 0.0036 |

O = Orkney, SR = South Ronaldsay, W = Westray, Lower = Lower 95% CI, Upper = Upper 95% CI, SD = Standard deviation, SE = Standard error.

Parametric and non-parametric MANOVAs showed significant differences among mandible groups (Wilks' $\lambda$ = 0.0006, $F_{352, 2421}$ = 7.0860, $p < 0.0001$; non-parametric MANOVA F = 8.2320, $p < 0.0001$, Appendix 11) and skull groups (Wilks' $\lambda$ = 0.0006, $F_{306, 1513}$ = 6.4670, $p < 0.0001$; non-parametric MANOVA F = 10.25, $p < 0.0001$, Appendix 12).

For the mandible and skull groups, there were no correlations between Procrustes and genetic distances (for mandibles: $R = -0.0823$, $p = 0.6803$; for skulls $R = 0.0617$, $p = 0.3688$), but there were significant correlations between geographic and Procrustes distances (for mandibles: $R = 0.3568$, $p = 0.0363$; for skulls $R = 0.6855$, $p = 0.0046$).

The classification of mandible and skull samples into their respective groups with the DA gave similar results, with 86.8% mandibles correctly classified (69.5% with the cross-validation method) and 84.6% skulls correctly classified (65.4% with the cross-validation method). The CVAs showed that the Continental and island groups were somewhat different (Fig. 4.5A, B). The groups from Navarra for the

mandible and skull datasets were easily distinguishable from the Ireland and Orkney islands groups, while the mandible samples from Belle Île clustered within the range of other continental samples and island groups. The two localities from Westray (Ness and Loch of Swartmill) showed pronounced differences and clustered separately from other island groups in the mandible and skull datasets, similarly to what was found in Chapter 3. Other island groups and localities showed no distinction, either in the mandible or skull datasets.

Dividing the mandible and skull datasets by island size categories gave a very different perspective on shape variation. In general, there was a higher classification percentage: 84.2% of mandibles were correctly classified (74.6% with the cross-validation method), and 93.9% of skulls were correctly classified (86.4% with the cross-validation method) to their original island size categories. The CVA plots showed a clear discrimination of island size categories for the mandible and skull datasets (Fig. 4.6A, B). The continental category was easily distinguishable from all the other categories, the large island size category was separated from medium and small island size categories, but the medium and small island size categories were somewhat mixed. The differentiation by island size category was more evident with the skull than with the mandible dataset.
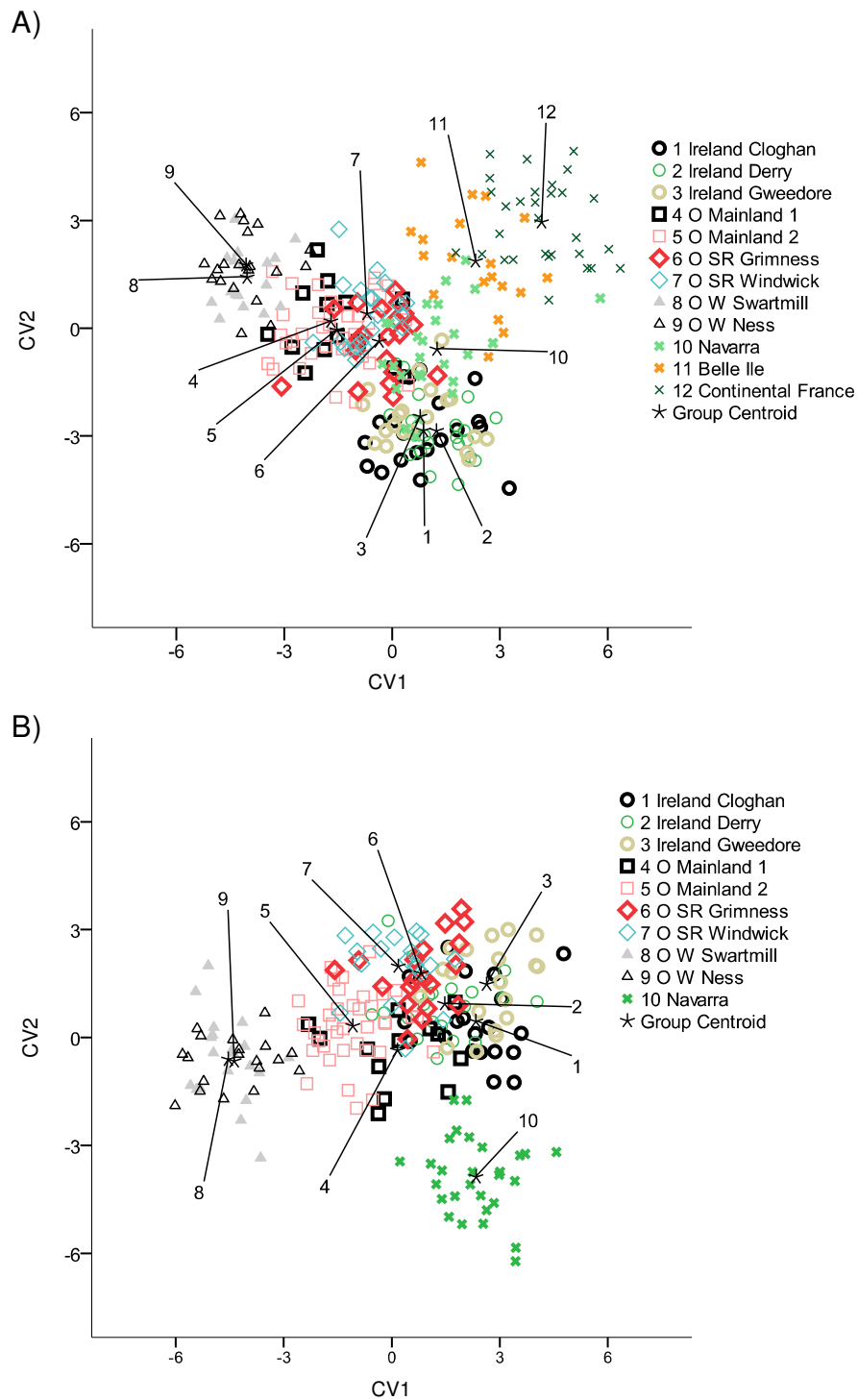
**Figure 4.5.** Canonical Variate Analysis of shape variables of A) mandibles and B) skulls of *Sorex minutus* from the Pyrenean lineage showing group differences. The first two canonical variates (CV) were used to describe variation among groups. O = Orkney, SR = South Ronaldsay, W = Westray.
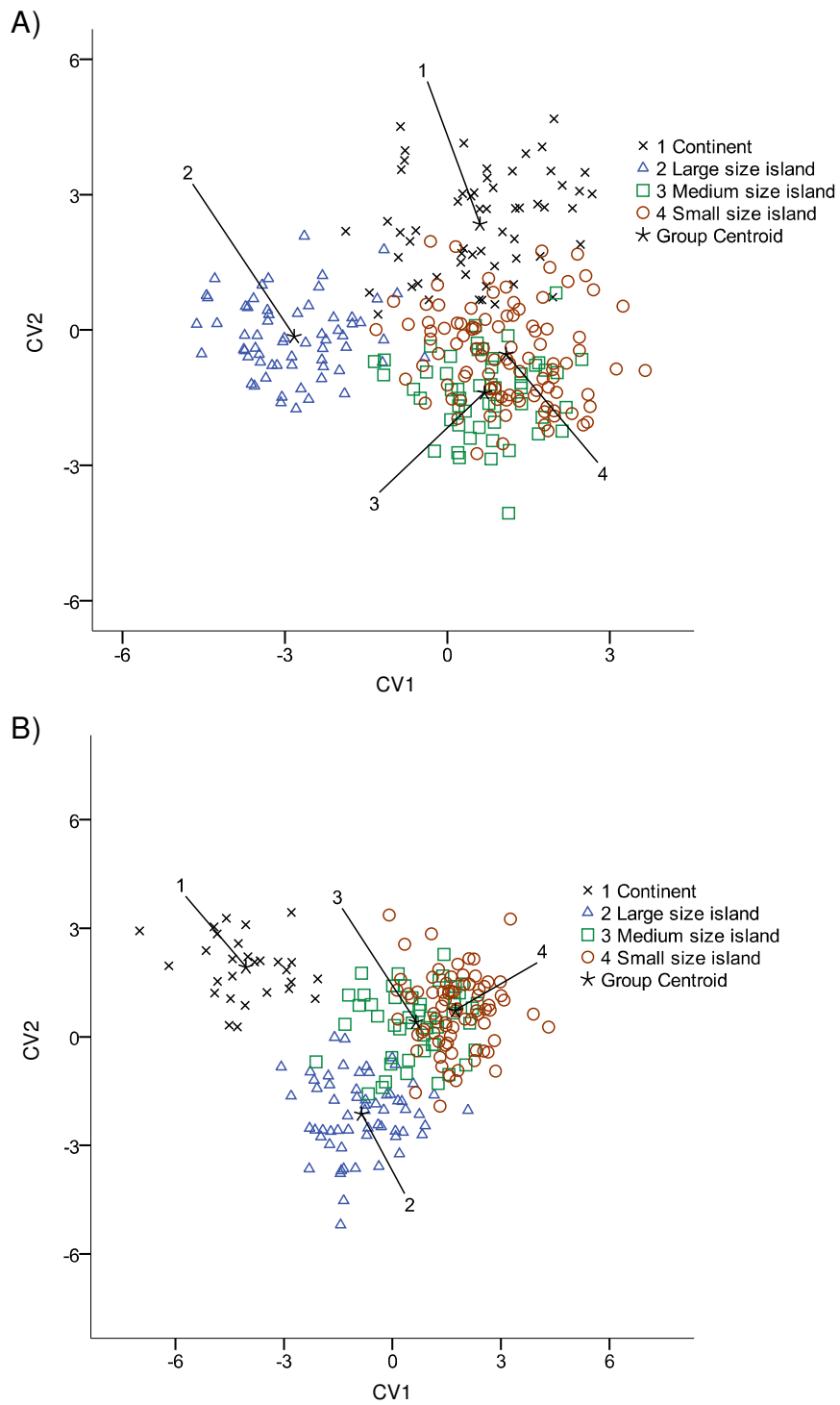
**Figure 4.6.** Canonical Variate Analysis of shape variables of A) mandibles and B) skulls of *Sorex minutus* from the Pyrenean lineage showing island size differences. The first two canonical variates (CV) were used to describe variations among groups. Continent = Continental European samples, Large size island = Ireland, Medium size island = Orkney Mainland, Small size island = South Ronaldsay, Westray and Belle Île.

## 4.4 Discussion and conclusions

Within the Pyrenean lineage, the Continental, Ireland, Orkney islands and Celtic fringe groups are closely related and share an intricate colonisation and migration history; however, they are geographically separated and currently isolated, aspects that could have affected their genetic and morphological diversity, and which I explored here.

The migration rates between the Continental and Celtic fringe groups were almost equal, showing a slight tendency of the former as a source area for the later. The phylogenetic network also supported a migration route from continental Europe to Britain, where the most internal haplotype for the full network was found in the Pyrenean group and some Celtic fringe haplotypes expanded from it. This is as expected; at the LGM, Britain would not have been occupied by pygmy shrews (Yalden 1982, 2007) and colonisation would have been from continental Europe over a landbridge that existed until 8.5 KYA (Weninger et al. 2008).

It is interesting that all samples from the Orkney islands also belong to a Celtic fringe described by Searle et al. (2009). The Celtic fringe is the vernacular name given for the culturally and genetically distinct human occupants of the western and northern periphery of Britain, reflecting the scenario that the Celts spread and occupied these regions 2 – 3 KYA and replaced earlier inhabitants (Forster and Toth 2003, Forster et al. 2004, Oppenheimer 2006, Searle et al. 2009). Other small mammals such as the common shrew *Sorex araneus*, bank voles *Myodes glareolus*, water voles *Arvicola terrestris* and field voles *Microtus agrestis*, show a similar peripheral distribution for genetic lineages; however, these species colonised the British Isles soon after the LGM and arrived there before the Celts (Searle et al. 2009).

The presence of an identical haplotype in Northern Spain (Navarra, close to the Pyrenees) and Ireland suggests a direct and long distance human-mediated migration event of *S. minutus* across the sea. However, it is possible that this haplotype was more widespread in continental Europe or it was present in a hypothetical population occupying the landmass between continental Europe and Britain, in which case the migration event would have been across a shorter distance. In the Eurasian phylogeographic study (Chapter 2), I found a few cases of identical haplotypes shared by individuals from distant localities, for example in Britain and the Netherlands (in the North-Central European lineage) or in Serbia and Macedonia (in the Balkan lineage); therefore, it is possible that shared haplotypes of the Pyrenean lineage between continental Europe and Britain exist.

The Celtic fringe hypothesis considers a two-phase colonisation pattern, with one genetic type colonising mainland Britain, then taken by humans to many of the other surrounding islands, but later partially replaced on mainland Britain by a second invading genetic type arriving from the east, generating the peripheral distribution of the first type in the British Isles (Searle et al. 2009). The genetic similarity of the haplotypes of *S. minutus* from South-Western Europe and those from Ireland, the Orkney islands and other Celtic fringe populations can be explained as a first colonisation wave of the 'Pyrenean' lineage from the continent into the British Isles, while the widespread distribution of *S. minutus* in mainland Britain and the genetic similarity with the 'North-Central European' lineage correspond to a second colonisation wave into Britain.

The low genetic diversity found in Ireland and the Orkney islands reflects colonisation of these islands by a restricted number of founders. Similarly, low mean cyt *b* nucleotide diversity values for *S. araneus* are present on islands of the Inner Hebrides off the west coast of Scotland (mean = 0.00105), with nucleotide diversity values ranging between 0 and 0.00316 (White and Searle 2008a); however, the

overall genetic diversity levels there were larger than the ones found in this study. Because Ireland and the Orkney islands have apparently not been connected to mainland Britain during low sea level and natural colonisation to these islands is unlikely (Yalden 1982, Stuart 1995, Lambeck and Purcell 2001), human-mediated colonisation could explain the presence of *S. minutus* on those islands, most likely transported accidentally during the Neolithic. Archaeological evidence indicates that the Neolithic people on Orkney could have easily travelled between islands, and that transportation by sea was easy and more convenient than travelling by land through the rugged interior (Phillips 2004). Spanish and French populations of the common vole *M. arvalis* are closely related to the Orkney vole, showing a comparable pattern to *S. minutus*, so an introduction by Neolithic people from South-Western Europe onto Orkney is plausible, and the high nucleotide diversity there even suggests multiple introductions (Haynes et al. 2003), but rodents may survive longer than shrews on a boat.

The high estimated migration rate between the Celtic fringe and Orkney islands groups indicates that the Orkney pygmy shrew likely originated from a source population in Northern Scotland; however, I did not find shared haplotypes between Orkney and the Celtic fringe or the continent, but most haplotypes from Orkney were directly connected to Celtic fringe haplotypes by only one mutational step, and not to continental haplotypes. Similarly, *S. araneus* cyt *b* haplotypes found on the islands of the Inner Hebrides were generally not found on the Scottish mainland, but most appear to have arisen from one of the haplotypes found on the mainland (White and Searle 2008a). Further sampling along the shores of Northern Scotland may provide haplotypes of *S. minutus* identical to those from the Orkney islands.

It is unlikely that the colonisation of Ireland and Orkney by *S. minutus* was more recent, in historical times (i.e., in the last hundreds of years), because the

island populations show some level of divergence in the cyt *b* gene. Moreover, the AMOVAs indicate that the little genetic variability in the sample was significantly structured among the islands and the Pyrenean group, and within the islands. However, the significant morphological variation observed here, mainly between islands and continental groups, cannot be used to distinguish a recent from an old colonisation event because morphological traits in mammals can evolve quickly on islands, in a matter of a few decades after colonisation (Pergams and Ashley 2001, Millien 2006).

Mandibles and skulls from the Orkney islands did not show the latitudinal trend observed with the full dataset (discussed in Chapter 3), as expected from a dataset more restricted latitudinally and closely related. Both datasets also showed distinct patterns of size variation. Most of the significant differences of mandible CS involved comparisons with the continental populations, while the only significant differences of skull CS involved comparisons with the Westray populations.

The multivariate analysis on shape showed significant differences for mandibles and skulls. There were also some apparent shape differences in the CVA when I divided the sample into groups based on phylogeny and/or geographical origin: Belle Île and continental France, and the Westray groups were differentiated from the other samples from Navarra, Ireland and Orkney in terms of mandible shape. The CVA on skull shape showed that Westray and Navarra were both distinct and separate from other groups from Orkney and Ireland. Unfortunately, I did not have skulls samples from Belle Île and continental France to compare with the mandible pattern of shape variation. However, these results indicated a similar pattern of shape variation for both mandibles and skulls overall, with Westray and Continental samples distinct from samples from Ireland and other Orkney islands.

Island size was a better discriminator of shape variation, because there were marked shape differences when I divided the mandible and skull datasets into island

size categories. For the mandible and skull datasets there were similar patterns of shape variation: The Continental and Large island groups were distinct from each other and from the Small and Medium sized islands, while the Small and Medium sized islands were very similar to each other. Westray, a small size island, had large pairwise Procrustes distances from other islands and the continent, and grouped separately in the CVA. This island also had zero nucleotide and haplotype diversity (only one haplotype was found), although a large number of individuals were sequenced, and it shared its haplotype with Orkney Mainland, likely presenting inbreeding effects.

Populations on small islands are usually smaller and more isolated than populations on larger islands or in the mainland, making them susceptible to extinction and stochastic demographic processes affecting genetic diversity (Frankham 1998), which can have an effect on the developmental stability and fluctuating asymmetry. In *S. araneus*, the significant correlation between heterozygosity and fluctuating asymmetry among the Inner Hebrides Islands were driven by the effects of the smallest island in that study (White and Searle 2008b). In this study, however, I did not assess the fluctuating asymmetry of the Orkney pygmy shrew, but I would expect to find the strongest effects on smaller islands with smaller genetic diversity values.

The results shown and discussed here represent the first analysis of genetic and morphological diversity of *S. minutus* on the Orkney islands, and it is the first study of morphological variation within Ireland in comparison to continental populations.

# Chapter 5

# General discussion and conclusions

## 5.1 Phylogeographic structure of *Sorex minutus* in Europe

It has long been recognized that the climatic oscillations during the Quaternary had profound effects on species ranges and the process of speciation (Haffer 1969, Hewitt 1996, Davis and Shaw 2001). The range shifts varied with latitude and topography, depended on individual species responses of migration and adaptations, and resulted in demographic changes that affected the genetic diversity and the geographical distribution of the genealogical lineages (i.e., the phylogeographic structure) of species (Hewitt 2000, 2004). Quaternary refugia represent the geographical regions that species inhabit during periods of glacial and interglacial cycles when there is the maximum contraction in geographical range (Stewart et al. 2010). Traditionally, it has been hypothesised that boreal and temperate species were restricted to Mediterranean refuge areas during glacial maxima in the Southern European peninsulas of Iberia, Italy and the Balkans. This glacial refugia hypothesis is supported by palynological studies, but also by phylogeographical studies on the meadow grasshopper *Chorthippus parallelus*, hedgehog *Erinaceus europeus/concolor* and brown bear *Ursus arctos*, which became the three paradigms to explain the colonisations of northern regions from Southern Europe by temperate species (Taberlet et al. 1998, Hewitt 2000). More recently, palaeontological, palynological, phylogeographic and species distribution modelling indicate that Quaternary refugia in Europe were more complex than currently believed with some temperate refugia further north (Bilton et al. 1998, Stewart and Lister 2001, Kotlík et al. 2006, Provan and Bennett 2008, Fløjgaard 2009, Vega et al. 2010b).

In this study, the phylogeographic analysis on the Eurasian pygmy shrew *Sorex minutus* revealed six cytochrome (cyt) *b* lineages, including an Iberian, North-Central Italian, South Italian, Balkan, Pyrenean and North-Central European, which apparently showed allopatric evolution in southern and northern glacial refugia, followed by population expansions and the establishment of contact zones. The genetic variability of this mitochondrial gene makes it a useful molecular marker for phylogenetic analysis and has been the marker of choice due to the lack of recombination, putative neutrality, smaller effective population size and shorter time for reciprocal monophyly between populations (Hickerson et al. 2010).

The cyt *b* genetic lineages found in the Mediterranean peninsulas are geographically restricted and did not contribute to the northward colonisation of Europe, a pattern first described by Bilton et al. (1998) but later criticised by Michaux et al. (2003) because of the restricted number of samples. Here, I corroborate previous findings on *S. minutus* with a more extensive sampling and comprehensive analysis, and show that, at least for this species, the Southern European peninsulas were refugial areas, while Northern and Central Europe were colonised by expanding populations located further north than the traditional southern glacial refugia.

The high genetic and haplotype diversities found within all the cyt *b* lineages, whether the purely northern lineages or the relatively limited southern forms, indicate that populations have not suffered recent genetic bottlenecks, further supporting the hypothesis of both southern and northern glacial refugia in Europe. Other phylogeographic studies have shown a similar pattern for small mammals, but *S. minutus* is one of the few species that is widely distributed in the three Mediterranean peninsulas, and Central and Northern Europe; therefore, it is an excellent model for understanding the effects of the glaciations in Europe and the colonisation history during the Pleistocene and the postglacial.

Considering the biogeographical traits of *S. minutus*, namely cold tolerance, present-day northerly distribution, habitat generalist, short generation times and small body size (Bhagwat and Willis 2008), it is possible that *S. minutus* populations were more widely dispersed during colder times than currently appreciated. Several cold tolerant species, such as the arctic fox, lemmings, reindeer, ptarmigan, etc., had larger geographical ranges during the glacials than during interglacials (Stewart and Dalén 2008, Stewart et al. 2010), and while this is probably not true for *S. minutus*, the species need not have been as restricted as implied by the grasshopper/hedgehog/bear paradigm for temperate species. Ecological niche models for *S. minutus* and other small mammals for which northern refugia have been hypothesised show relatively widespread distributions at the Last Glacial Maximum (LGM), mostly restricted to the north by the ice sheets, so it is possible that temperate but cold tolerant species were able maintain a wide range in Central Europe during the glacial periods (Fløjgaard 2009, Vega et al. 2010b). Moreover, *S. minutus* is currently found at higher altitudes in the Mediterranean peninsulas (usually higher than 500 m.a.s.l; Madureira and Magalhães 1980, MacDonald and Barrett 1993), likely because the high temperatures in the lowlands does not suit it (Taylor 1998), which suggests that in the present time populations there could be in interglacial refugia (Hilbert et al. 2007, Bennett and Provan 2008, Stewart et al. 2010).

For the nuclear markers (two concatenated Y-chromosome introns and the BRCA1 autosomal gene) there was very low sequence polymorphism resulting in phylogenetic reconstructions with low power and poor resolution, so further phylogeographic inferences with these markers were not performed and could not be compared with the mitochondrial (mt) DNA findings. Nonetheless, the use of molecular markers with different modes of inheritance provides a better chance to reconstruct the evolutionary history, and it also avoids the over-interpretation of

results with single trees, which can be affected by stochastic processes (Brito and Edwards 2009).

The concatenated Y-chromosome introns showed equivalent clades to the cyt *b* phylogroups, with distinct North-Central Italian, South Italian and Southern Balkan groups, plus a clade including all Pyrenean and Iberian samples, and a clade including all North-Central European, Northern Balkan and some samples from the northernmost distribution of the North-Central Italian phylogroup. This result provided additional evidence for a geographical structure in Europe below the species level of *S. minutus*. The differences between the cyt *b* and concatenated Y-chromosome introns trees, and the clustering of the Pyrenean with Iberian, Balkan with North-Central European and North-Central European with North-Central Italian clades could be due to sex biased dispersal near the contact zones (Prugnolle and de Meeus 2002); however, I consider that this is more likely the result of the lower genetic polymorphism and, thus, lower phylogenetic resolution of the Y-chromosome introns compared to the cyt *b*.

The BRCA1 nuclear gene had very little genetic variation and did not resolve distinct phylogroups, except for a South Italian and a Pyrenean plus Iberian clades. This result was not unexpected because other studies have also shown that the number of informative sites in nuclear genes is small and the phylogenetic resolution is poor, a problem that can be resolved by concatenating multiple genes together (Brito and Edwards 2009).

## 5.2 Historical demography and colonisation routes

One important result in this study was the finding that Northern and Central Europe, including the British Isles, have been colonised by two lineages, the Pyrenean and the North-Central European. These two lineages show a characteristic signature of a population expansion, strongly supported by the

137

mismatch distributions, the significant neutrality and $R_2$ tests, and the star-like pattern of the haplotype network (Fig. 2.3). These results were also characteristic of continental populations only, i.e., excluding the populations on the British Isles. Furthermore, the North-Central European lineage had the most widespread geographical distribution and showed a subtle geographical differentiation among Eastern Europe, Central Europe and Britain, which argues in favour of relatively isolated populations; however, a more detailed analysis with a faster evolving molecular marker would provide a better understanding of the population genetic structure of the North-Central European phylogroup.

The North-Central Italian lineage also showed a significant population expansion signature, likely a result of the colonisation of previously glaciated areas in the Alps, and forms a long contact zone with the Pyrenean, North-Central European and Balkan lineages in Central Europe. The Balkan lineage, in contrast, showed a stable population size through time and no evidence of a recent population expansion. It is possible that in the Balkans, populations were able to maintain relatively constant size in the heterogenic landscape there, but it is also possible that the multimodal mismatch distribution is a result of mixing very distinct populations into a single group. Actually, the cyt *b* phylogenetic trees showed two distinct and divergent sub-lineages in the Balkan peninsula that deserve further attention (see Fig. 2.2). Martino's vole (Dinaric snow vole) *Dinaromys bogdanovi* is a Balkan endemic genus and species with three highly divergent lineages in the North-Western, Central and South-Eastern Balkans, with the first two lineages showing multimodal mismatch distributions, supporting a multirefugia scenario in the Balkans and stable population sizes during the glaciations (Kryštufek et al. 2007, Bužan et al. in press). In Iberia, however, the lack of a significant population expansion signature for *S. minutus* is likely a result of low sample size. Nonetheless, the Iberian peninsula is considered an important and complex refugial

area with several microrefugia and endemics for temperate species of fish, amphibians, reptiles, mammals, invertebrates and plants (Gómez and Lunt 2007). Future studies on *S. minutus* from the Iberian peninsula might be able to show significantly differentiated populations and independent demographic histories in the Sistema Central and Cantabrian mountain ranges.

Searle et al. (2009) described a Celtic fringe in the British Isles composed of samples from the western periphery of Britain, which in this study belonged to the Pyrenean lineage. They hypothesised a selective sweep (or recent adaptive substitution) coming from the east to explain the western and northern peripheral distribution of the Celtic fringe samples and the widespread distribution of the North-Central European phylogroup in mainland Britain. The patterns of an expanding population and a selective sweep can be similar and difficult to distinguish from each other, because the synonymous and segregating sites in the cyt *b* gene used in the population expansion tests can be genetically linked to sites under selection (Ballard and Whitlock 2004). For example, a mtDNA haplotype, characterized with a non-synonymous mutation resulting in a transition between Leucine/Serine aminoacids of the protein encoded by the ATP6 gene, has been implicated in the non-shivering thermogenesis and differences in respiratory and winter survival of the white-toothed shrew *Crocidura russula* (Fontanillas et al. 2005). Similarly, it is possible that natural selection favoured individuals of the North-Central European cyt *b* lineage on Britain, replacing the Pyrenean lineage and originating the Celtic fringe. If this is so, the population expansion pattern observed there could have been a result of a selective sweep. However, the colonisation and spread of *S. minutus* over favourable habitat in Britain most likely produced a population expansion pattern that could have been amplified by a selective sweep effect. There were very few non-synonymous substitutions in the cyt *b* gene, no evidence of selection based on the McDonald-Kreitman test and no violation of the neutrality

hypothesis compared to an external group, so it is possible that within *S. minutus* there are no selection forces acting on the cyt *b* gene. Unfortunately, this test is not suitable for detecting very recent selective sweeps because non-synonymous and synonymous mutations, linked to the beneficial mutation, will be similarly affected by the selective sweep (Nielsen 2005), and a significant signature could also result from a combination of many slightly deleterious non-synonymous mutations and population expansions (McDonald and Krietman 1991).

Contrary to most phylogeographic studies in Europe, the timings of the population expansions in *S. minutus* pre-date the LGM. Other studies have also shown population expansion events pre-dating the LGM, for example for the weasel *Mustela nivalis* in the Western-Palaearctic region (Lebarbenchon et al. in press) and for the Asian clade of the *Crocidura suaveolens* group (Dubey et al. 2006). However, the timing of the population expansion was calculated assuming a molecular clock and a standard 2% substitution rate per million years, which could give problematic results. If I had used a higher substitution rate as molecular clock, the divergence and population expansion times would have been more recent. Different substitution rates for small mammals and shrews are commonly used in phylogeographic studies, for example Ratkiewicz et al. (2002) used a mutation rate of 2.5% per million years for studying the evolutionary history of karyotypic groups of the common shrew *S. araneus* (but originally used for cichlid fish); Haynes et al. (2003) used a 6-10% mutation rate for *Microtus arvalis*; and Yannic et al. (2008) used a divergence rate of 7.1% per million years with a confidence interval between 5.7 and 8.5% based on Maximum Likelihood distances for studying the evolutionary history of the western *S. araneus* group. Another problem with the molecular clock is known as the exponential decay law, which states that the divergence and population expansion times of a lineage can be overestimated when making the erroneous assumption of a linear relationship between genetic differences and time

(Ho et al. 2005, Ho and Larson 2006); but Bandelt (2008) pointed out that there are problems with the way the decay law was modelled originally. Therefore, the timing of events in this and other studies have to be taken with caution because they depend on the chosen mutation rates, but also on other factors such as the time of the fossil calibration for Bayesian estimates of TMRCA, the number of generations per year assumed for the species under study (accurate in the case of the pygmy shrew, Corbet and Harris 1991) and the rate heterogeneity among lineages.

## 5.3 An emerging pattern of northern glacial refugia

The comparison of the results obtained here with those elsewhere shows an emerging pattern of glacial refugia in Mediterranean peninsulas and further north in Central Europe for several species (Fig. 5.1A–D).
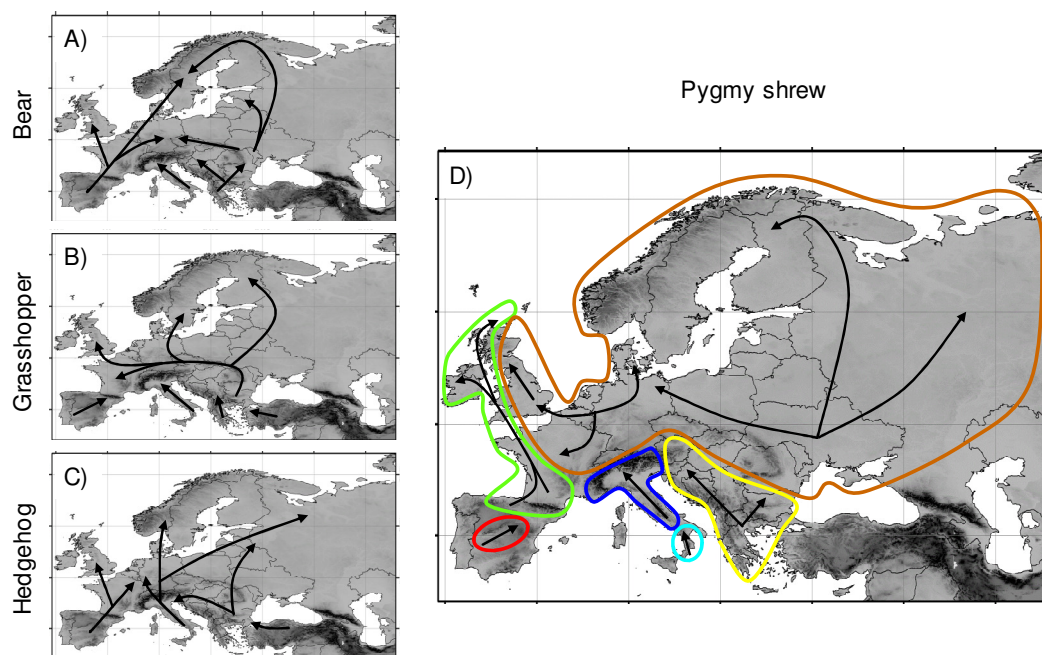


**Figure 5.1.** Three paradigms of postglacial colonisation routes from Southern Europe: A) the bear, B) the grasshopper and C) the hedgehog (Hewitt 2000). The addition of D) the pygmy shrew gives a more complete picture and exemplifies an emerging pattern of northern refugia. Arrows represent general colonisation routes. Coloured areas are an approximation of the geographical distribution of the cytochrome *b* lineages obtained in this study (see Fig. 2.2 for lineages).

The bank vole *Myodes* (*Clethrionomys*) *glareolus* has a phylogeographic pattern very similar to that found with *S. minutus*: there are three Mediterranean (Iberian, Italian and Balkan) and three northern phylogroups (Eastern, Western and the Ural Mountains) significantly differentiated, with heterogeneous mismatch distributions and population expansion signatures in the northern phylogroups but long-term stability for the Mediterranean populations, and most of the genetic structure distributed among lineages (Deffontaine et al. 2005). Furthermore, the bank vole has a northern glacial refugium in the Carpathians from where it expanded at the end of the last glaciation and from where it recolonised northern latitudes (Kotlík et al. 2006, Wójcik et al. 2010). The field vole *Microtus agrestis* has three lineages that derived from separate refugia, one or two in the Iberian peninsula, one in Western Europe and another one in Eastern Europe (Jaarola and Searle 2002). The Iberian and Western lineages have a contact zone in Southern France and Switzerland, while the Western and Eastern lineages have a contact zone in Fennoscandia, likely originating from different colonisation routes (Jaarola and Searle 2002). The common vole *M. arvalis* has significantly differentiated lineages in the Iberian and Italian peninsulas, distinct from the lineages further north in Western, Central and Eastern Europe, which indicate a limited effect of southern refugia in shaping the current patterns of genetic diversity (Haynes et al. 2003, Heckel et al. 2005). Interestingly, the *Sorex araneus* group also shows divergence of Southern Mediterranean populations from Central and Eastern European groups, with *S. granarius* in Iberia, *S. coronatus* in the Pyrenees, France, Western Germany and Switzerland, *S. antinorii* in Italy and other *S. araneus* chromosomal races in Central and Eastern Europe, plus *S. samniticus* (endemic to Southern Italy) as a sister species (Yannic et al. 2008). Moreover, several mammal species show a genetic distinction in Southern Italy, including *M. glareolus* (Amori et al. 2008), *M.*

*brachycercus* (Castiglia et al. 2008), *Talpa romana* (Ungaro et al. 2001), *Lepus corsicanus* (Pierpaoli et al. 1999) and *Sciurus vulgaris* (Grill et al. 2009), plus other vertebrate species, denoting the importance of the Calabrian arc as a hotspot for European biodiversity (Vega et al. 2010a).

**5.4 Natural and human-mediated colonisation**

The widespread European distribution of *S. minutus* is a mixture of natural and human-mediated colonisation events. The fossil record of *S. minutus* suggests an old and natural colonisation of the European continent since the Early Pleistocene, originating from the East. The current genetic structure suggests more recent divergence events leading to a variety of phylogroups that attained their present distribution through natural colonisation and (in parts of the British Isles) through human introductions.

The cyt *b* analysis indicates that Britain was colonised naturally by two distinct lineages, the Pyrenean and the North-Central European. Sea level changes during the LGM indicate the presence of a landbridge between mainland Europe and Britain (Yalden 1982, Stuart 1995), which could have been a temporal and suitable land corridor for *S. minutus* and allowed this and other species to colonise Britain by their own means (Yalden 1982). The genetic similarity of continental European and British samples supports this view. The presence of *S. minutus* and other small mammals in Ireland, however, is puzzling and has been the focus of much research.

The close genetic relationship of the haplotypes from the Orkney islands, Ireland and other Celtic fringe samples from Britain with South-Western Europe suggests a geographical origin there. In particular, the occurrence of exactly the same haplotype in Northern Spain and Ireland (where it is the most common haplotype) is striking. However, the direct colonisation of Ireland from Northern

Spain with people seems unlikely given the geographical distance (about 1000 Km); therefore, it seems more plausible that populations in Britain sharing the haplotype now found in Northern Spain were the source for the Irish lineage. It should be noted that haplotypes separated by only one mutation from haplotypes in Northern Spain have been found in France and Britain (Fig. 4.2), and it would not be surprising if some Pyrenean lineage haplotypes identical to haplotypes currently found in Spain managed to colonise Britain and subsequently make their way to Ireland. It is likely that it was humans that transported the shrews from Britain to Ireland, given the lack of evidence of a landbridge (Searle et al. 2009). In support of this model, it is notable that there is currently a wide diversity of haplotypes in the Celtic fringe (Fig. 4.2), which could easily have included the haplotype that started off the colonisation of Ireland. Further sampling of the Celtic fringe on Britain could shed more light onto this. There is particularly poor representation of the Pyrenean lineage in Wales and South-Western England (Fig. 4.1), and it may be that this was the source area of the Irish colonisation. I have provided strong evidence that the Orkney pygmy shrew originated from populations in mainland Britain belonging to the Pyrenean lineage, detected by the genetic similarity to other Celtic fringe haplotypes and by the migration analysis. The pygmy shrew was most likely taken to the Orkney islands by boat during the Neolithic through trade routes among nearby coastal areas widely established by that time (Phillips 2004, Zvelebil 2006) rather than by natural processes like rafting, swimming or migration over ice. The closest island of the archipelago to the mainland (South Ronaldsay) is about 10 Km away from the coast, and the Orkney islands were not connected to the mainland during lower sea levels in the postglacial (Yalden 1982). Furthermore, the pygmy shrew from Westray seems to have originated directly from Orkney Mainland, about 16 Km away, because Westray has only one haplotype shared with Orkney Mainland. Human-mediated colonisation of the islands surrounding Britain and

Ireland does not seem implausible, because other species of small mammals on islands in the Mediterranean, Canary Islands, Madeira archipelago and the British Isles were also transported there by humans, including house mice, white-toothed shrews (*Crocidura* spp.) and *A. sylvaticus* (Delany and Healy 1966, Gündüz et al. 2001, Renaud and Michaux 2007, Michaux et al. 2007).

For several islands separated by short distances or close to the coast, it is possible that *S. minutus* colonised by its own means. At least for *S. araneus* and *S. minutus*, natural island colonisation less than 500 m from the mainland is possible in the cold waters of Lake Koitere in Finland (Hanski 1986). Hoy and Orkney Mainland share the same haplotypes and are geographically proximal, so it is possible that the two islands maintained gene flow in the past when the sea level was lower. Similarly, White and Searle (2008a) hypothesised that Jura and Islay remained in contact until recently, and that gene flow still occurs at a low rate between mainland populations and Skye and Seil in the Inner Hebrides given the high level of microsatellite diversity and low genetic differentiation. Moreover, the haplotypes from South Ronaldsay likely originated from a different source population in Northern Scotland, because they are different from the haplotypes found on the other Orkney islands. Further studies with microsatellites would provide a better understanding of gene flow between the mainland and the Orkney islands.

## 5.5 Morphological diversity of *Sorex minutus* in Europe

Morphological differentiation of *S. minutus* in Europe is much less evident than genetic differentiation. Without an a priori grouping of the mandible and skull samples, there were no consistent shape differences between samples in the Relative Warp (RW) analyses; however, dividing the mandible and skull samples into the phylogroups detected with cyt *b* revealed small but significant shape variation among the groups with the Canonical Variates Analysis. Multivariate

analyses of variance detected morphologically different phylogroups, but the shape differences did not recover the phylogenetic relationships among them. There are several stochastic, developmental, environmental and genetic factors affecting mandible and skull structures (Klingenberg 2001, Monteiro et al. 2003, Caumul and Polly 2005, Cardini et al. 2007); therefore, it is not surprising that the genetic relationships were not properly recovered. Other studies have shown disagreement between morphological and phylogenetic findings. For example, a three-dimensional geometric morphometric study of *Arvicanthis* (an African murid) showed a close relationship of morphology with environmental background, but only a partial agreement with phylogeny, possibly as a result of repeated, convergent evolution of similar morphology and adaptation to local climatic conditions (Fadda and Corti 2001); and the trends in shape variation of spiny rats of the genus *Trichomys* disagreed with the expected phylogenetic pattern (Monteiro and dos Reis 2005).

There are few detailed intraspecific geometric morphometric studies of wild mammals, although geometric morphometrics is a good method for studying shape variation below the species level (Loy 1996). Inter- and intraspecific studies have shown that marmots, *Trichomys apereoides*, vervet monkeys, phyllostomid bats, etc. have notable shape differences, which are usually correlated with environment, geography, phylogeny and feeding habits (Corti et al. 2001, Cardini 2003, Caumul and Polly 2005, Monteiro et al. 2003, Cardini and O'Higgins 2004, Nogueira et al. 2009). *Sorex minutus* is considered a food generalist because of the high diversity of prey it feeds on. Thus, it is possible that the shape of mandibles and skulls of *S. minutus* responds weakly to habitat changes, because the selection pressures on anatomical structures involved in mastication might be very similar and there is no prey specialization in different geographical regions (Castién and Gosálbez 1999).

Size, however, showed a marked latitudinal trend on the pygmy shrew mandible across Europe with an increasing size with decreasing latitude, contrary to Bergmann's rule. It is possible that prey in southern latitudes have harder exoskeletons, so selection pressures could have favoured bigger mandibles and a stronger bite force, without affecting overall shape, as shown for South Italian pygmy shrews (Vega et al. 2010a). There was a strong allometric effect on mandibles, so most of the shape variation observed among continental samples was strongly correlated with size and latitude. The skull sample showed a weak but still significant latitudinal pattern of size, but the skull is less involved in the mastication process than the mandible. Ochocińska and Taylor (2003) found a stronger latitudinal pattern on *S. minutus* skulls than in this study, using samples from across Eurasia but with a smaller sample size per locality. They did not use a geometric morphometrics approach; instead they measured the size of the skull as the distance from the condylus occipitalis to the front edge of the praemaxillare, a good indicator of overall body size. Nevertheless, the interpretation of their results was equivalent to the one presented here for mandibles and skulls using a geometric morphometric approach and also support a latitudinal pattern.

## 5.6 Insights into island evolution of *Sorex minutus*

Besides the evident latitudinal trend of mandible and skull size and the little shape variation of continental samples, it seems that island evolution rather than phylogenetic differentiation in southern and northern refugia played an important role in morphological shape variation of *S. minutus*.

Insular populations of mammals appear to follow an 'island rule' (Van Valen 1973), which predicts an increase of body size for small mammals (gigantism) and a decrease of body size for large mammals (dwarfism) on island populations compared to mainland populations, probably as a result of limited insular resources,

147

less intraspecific competition and predation (Lomolino 2005). In addition, it is possible that morphological variation is a result of an 'island syndrome', traditionally described as the systematic differences in demography, reproduction, behaviour and morphology (including the island rule discussed previously) in island populations when compared to mainland populations (Adler and Levins 1994). These differences may evolve as a result of microevolutionary processes and after many generations of directional selection following colonisation and isolation, generating locally adapted island populations (Adler and Levins 1994). The effects of genetic variability on the morphological divergence of island populations and speciation have been widely explored in the past (Berry 1986); however, most of the factors involved in island differentiation depend on the genetic composition of the founding population (Berry 1996).

In this study, island pygmy shrews from Ireland and the Orkney archipelago were not bigger than continental populations, inconsistent with the island rule sensu Van Valen (1973), but the size of mandibles and skulls was consistent with the latitudinal pattern described for the full dataset. In contrast, populations in Ireland and the Orkney islands showed a marked shape distinction from continental counterparts, while continental samples showed an extensive overlap of shape variation as illustrated by the RW Analysis. Actually, samples of a genetically divergent *S. volnuchini* sample (with about 5% cyt *b* sequence divergence) showed little morphological variation from the continental groups of *S. minutus*.

It is interesting that the strong pattern of shape divergence between continent and islands evolved in just a few thousand years after the colonisation of Ireland and the Orkney islands. Mandible and skull shape from island populations was easily discriminated by island size and was easily distinguished from continental populations, regardless of the high genetic similarity. Founder effects, genetic bottlenecks, genetic drift due to small population size and geographical

isolation could have affected the morphological variation of *S. minutus* in Ireland and the Orkney islands. Samples from Westray were significantly different from other island populations in terms of mandible and skull size and shape, and the population there had the lowest genetic diversity values, whereas Orkney Mainland and Ireland populations had more heterogenous shape variation and higher genetic diversity values. The continental population from the Pyrenees, genetically related to the island populations analysed here, was significantly different in mandible and skull shape, had the most heterogeneous shape variation and had the highest genetic diversity values.

The shape and size variation on Westray and other small sized islands could be related to genetic bottlenecks and colonisation events from a low number of founders, while Orkney Mainland and Ireland, much larger islands, could have been colonised in several occasions and can maintain larger populations. In agreement with these results, other studies have also shown morphological divergence of recently colonised island populations. For example, the house mouse on the Canary Islands reflects some degree of differentiation and bigger molars than continental populations, which evolved in a matter of five centuries after the first introduction, although it is unclear if this is an adaptive or a random effect (Michaux et al. 2007). The mandible of the wood mouse *Apodemus sylvaticus* shows significant divergence on islands of the Mediterranean Sea correlated with remoteness from the continent and island size; moreover, there is a latitudinal mainland gradient in size and shape of mandibles and molars, but different individual patterns of differentiation on islands (Renaud and Michaux 2007). Apparently, the large-scale clinal pattern of morphological variation on the mainland populations of *A. sylvaticus* is favoured over adaptation to small-scale environmental variations, buffered by gene flow. Island populations, instead, show a mosaic pattern of evolution because

gene flow is disrupted due to geographic isolation and populations adapt differently to local environmental conditions (Renaud and Michaux 2007).

## 5.7 Conservation implications

One of the aims of this study was to provide a better understanding of European biodiversity, which could have implications for conservation. Although *S. minutus* is considered as a least concern species by the IUCN (IUCN 2008; http://www.iucnredlist.org/details/29667), the distinct lineages deserve more attention than this implies. Local and/or country based conservation efforts are highly valued (for example, on Britain and Ireland the pygmy shrew is protected by law), and genetic diversity is now considered an important aspect of global biodiversity (McNeely et al. 1990). Moreover, as the refugial areas in Southern Europe are often found in mountain ranges at the low-latitude margins of the present-day distribution ranges of species, they are most likely to contain rear-edge populations where selection for local adaptations could have resulted in the evolution of distinct ecotypes (Cook 1961, Hampe and Petit 2005). It has been argued that the populations that inhabit latitudinal margins deserve investigation and high conservation priority because they are important to determine the responses of species to modern climate change (Petit et al. 2003, Hampe & Petit, 2005). Clearly, the genetically divergent lineages of *S. minutus* found in this study should be considered as Evolutionarily Significant Units (ESUs), defined as genetically differentiated and reciprocally monophyletic lineages with a unique geographic distribution, which contain the evolutionary potential of the species (Moritz 1994).

## 5.8 Future directions

This study opens the door for further investigations. Population genetic studies in Mediterranean peninsulas and further sampling there would provide a

better understanding of the gene flow and structure of the genetic lineages, and would allow further investigation of the concept of 'refugia within refugia'.

In Southern Italy, it is possible that there are very distinct populations in the Calabrian arc. The Aspromonte and Serre Massifs in Calabria contain the southern-most populations of *S. minutus*, which were not described genetically or morphologically in this study, and could contain highly differentiated populations due to geographical isolation. Also, populations in Monte Pollino, Basilicata and in Catena Costiera, Calabria could hold the southern-most populations of the North-Central Italian lineage, separated from the South Italian lineage by the Crati Valley but relatively contiguous with the populations from the Apennines.

In the Balkans, it is likely that there are two highly divergent groups, as explained above. The high genetic divergence of this lineage indicates that it is one of the oldest lineages of *S. minutus* and highlights the importance of the Balkans as a biodiversity hotspot. Because of the historical, political, linguistic and topographic complexity of this peninsula, the scientific studies have accumulated slowly in comparison to other European regions and much remains to be investigated there (Kryštufek and Reed 2004).

Further sampling in Iberia would provide a better delimitation of the Iberian and Pyrenean lineages, and it is possible that both have a contact zone in the Cantabrian mountain range. I encountered one haplotype that clustered with the Balkan lineage and another one quite divergent from all the rest, which were eliminated from the analysis because of the uncertainty of locality data and contamination issues; however, this deserves more attention because they could represent relics of an old colonisation event. Regions in North-Western and Central Spain and Portugal have not been characterised in this or other studies, and samples from there could show other divergent lineages that could add to the refugia within refugia hypothesis of the Iberian peninsula.

Siberia and Eastern Europe were undersampled, and it would be interesting to obtain samples from there to study the phylogeographic structure. The highly divergent haplotypes from Ukraine, Russia and Poland could belong to a cryptic lineage in Siberia.

The mixing of *S. volnuchini* within the shape variation of *S. minutus* was unexpected, and it would be interesting to add other *Sorex* species to the geometric morphometrics analysis, as well as other Soricidae species with terrestrial, aquatic and fossorial habits. Instead of expecting to find a strong phylogenetic component of shape variation, I would expect to find an ecogeographical grouping of Soricidae species. Moreover, further sampling of *S. volnuchini* would provide a better understanding of the genetic relationship between these sister species.

Fluctuating asymmetry could be analysed in the context of different environmental correlates and population genetic diversity on the Orkney islands and other islands off the coast of Britain. There are several microsatellites developed for *Sorex*, and the genetic diversity values and asymmetry could be correlated with island size. For this, other islands of the Orkney archipelago should be sampled to give a more comprehensive picture of island evolution.

Finally, this is an interesting time for phylogeography and population genetics (Brito and Edwards 2009). The way molecular data is collected and analysed has been revolutionised by new genomic tools developed on model organisms and applied to evolutionary and ecological studies, research areas currently named as 'genomic phylogeography' and 'population genomics' (Brito and Edwards 2009). The transition from single to multiple loci in phylogeographic analyses is taking place fast. Larger datasets include more informative sites and increased nodal support, and more loci that sample whole genomes allow the inference of species trees rather than gene trees, and provide a more comprehensible understanding of population history and speciation processes.

Although I used cyt *b*, two Y-chromosome introns and an autosomal gene, this study is not different from a traditional phylogeographic study, because each molecular marker was analysed separately. mtDNA phylogeography has been essential for the development of the field, has provided the basic research and will not be abandoned soon given the relative simplicity of analysis and lower economic costs, and genome-informed studies will gain from traditional phylogeographic findings (Brito and Edwards 2009).

## 5.9 Conclusions

The Eurasian pygmy shrew *Sorex minutus* is a good model for understanding biological diversity, colonisation patterns and the effects of past climate change on biological diversity.

There is a mosaic of genetic lineages across continental Europe, characterised by different demographic histories and natural colonisation patterns, while island populations are characterised by recent natural and human-mediated colonisations. It is interesting that the pygmy shrew readily colonises islands as a transient stowaway of people despite its high metabolic requirements. Moreover, *S. minutus* also shows varying degrees of morphological variation. There is a latitudinal trend of size variation, contrary to one of the best known biogeographic rules, Bergmann's rule, with increasing size with decreasing latitude, a trend found to be common in other species of shrews. There is little shape variation among continental populations, but the differences are significant. Shape variation is strongly correlated with size, latitude and longitude, not with genetic divergence; therefore, environmental factors played a more important role than genetic differentiation in the morphological diversification of the species. Island evolution has been another factor strongly influencing shape variation, probably because of founder effects, genetic bottlenecks and inbreeding on small islands, but *S. minutus*

on islands does not seem to have been affected by an island syndrome of increasing body size.

This study has notably expanded previous findings on *S. minutus*, with a more precise statistical phylogeographic analysis of the genetic variability, colonisation routes, historical demography and structure across Europe. I have shown more clearly than before that the North-Central European lineage emerged from a northern glacial refugium, both through better characterisation of that lineage and better characterisation of populations in the Mediterranean peninsulas. A Pyrenean lineage is also very likely to have existed north of the traditional southern refugial areas. These studies are helping to develop *S. minutus* as a major model to study northern glacial refugia. *Sorex minutus* is not an easy species to obtain in large numbers, and the sampling described here represents a very substantial effort. However, it is a species that is unusually widespread and genetically subdivided and therefore can inform better than almost any other about the relative importance of southern and glacial refugia. This is important because the characteristics of species within a given area may reflect (in terms of genetic variability, adaptations) the nature of the refugial area from which they derived.

The study of island populations of *S. minutus* has not only been of interest with regards island evolution, but also with regards island colonisation. Although previous studies have investigated *S. minutus* in Ireland, my work has provided the level of sampling that has been necessary to develop a realistic colonisation scenario. It has also been possible to do the same for the Orkney islands. These studies add considerably to our understanding of how even prehistoric (Neolithic) humans can shape the genetic architecture of a species by moving particular forms around.

In summary, this has been (a) the first large-scale study of the morphological diversity of *S. minutus* across Europe with a geometric morphometrics approach, (b)

a large extension of previous studies considering the phylogeography of *S. minutus* across the same range, particularly focusing on the issue of northern versus southern refugia and human-mediated colonisation of the British Isles, and (c) the first study that has explored the colonisation, genetic variability and morphological diversity of *S. minutus* on small islands (the Orkney archipelago). The results presented here are of importance for the conservation of genetic and morphological diversity in one of the smallest mammals in Europe.

## References

**Adams DC, Rohlf FJ, Slice DE. 2004.** Geometric morphometrics: ten years of progress following the 'revolution'. *Italian Journal of Zoology* **71:** 5-16.

**Adler GH, Levins R. 1994.** The island syndrome in rodent populations. *The Quarterly Review of Biology* **69:** 473-490.

**Afzal-Rafii Z, Dodd RS. 2007.** Chloroplast DNA supports a hypothesis of glacial refugia over postglacial recolonization in disjunct populations of black pine (*Pinus nigra*) in Western Europe. *Molecular Ecology* **16:** 723-736.

**Agustí J, Blain H-A, Furió M, De Marfá R, Santos-Cubedo A. In Press.** The early Pleistocene small vertebrate succession from the Orce region (Guadix-Baza Basin, SE Spain) and its bearing on the first human occupation of Europe. *Quaternary International*.

**Amori G, Aloise G, Annesi F, Franceschini P, Colangelo P. 2008.** Filogeografia mitocondriale di Arvicola rossastra *Myodes glareolus* in Italia. *69° Congresso Unione Zoologica Italiana,* Senigallia, September 22–25. Abstracts, 18.

**Anderson MJ. 2001.** A new method for non-parametric multivariate analysis of variance. *Austral Ecology* **26:** 32-46.

**Ashton KG, Tracy MC, Queiroz AD. 2000.** Is Bergmann's rule valid for mammals? *The American Naturalist* **156:** 390-415.

**Avise JC 2000.** *Phylogeography. The History and formation of species.* Harvard University Press. Cambridge, USA.

**Avise JC, Arnold J, Ball RM, Bermingham E, Lamb T, Neigel JE, Reeb CA, Saunders NC. 1987.** Intraspecific phylogeography: the mitochondrial DNA bridge between population genetics and systematics. *Annual Review of Ecology and Systematics* **18:** 489-522.

**Ballard JWO, Whitlock MC. 2004.** The incomplete natural history of mitochondria. *Molecular Ecology* **13:** 729-744.

**Bandelt HJ. 2008.** Clock debate: when times are a-changin': Time dependency of molecular rate estimates: tempest in a teacup. *Heredity* **100:** 1-2.

**Bandelt HJ, Forster P, Rohl A. 1999.** Median-Joining networks for inferring intraspecific phylogenies. *Molecular Biology and Evolution* **16:** 37-48.

**Baquero RA, Tellería JL. 2001.** Species richness, rarity and endemicity of European mammals: a biogeographical approach. *Biodiversity and Conservation* **10:** 29-44.

**Barton NH, Hewitt GM. 1985.** Analysis of hybrid zones. *Annual Review of Ecology and Systematics* **16:** 113-148.

**Bennett KD, Provan J. 2008.** What do we mean by 'refugia'? *Quaternary Science Reviews* **27:** 2449-2455.

**Berry RJ. 1986.** Genetics of insular populations of mammals, with particular reference to differentiation and founder effects in British small mammals. *Biological Journal of the Linnean Society* **28:** 205-230.

**Berry RJ. 1996.** Small mammal differentiation on islands. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences* **351:** 753-764.

**Bhagwat SA, Willis KJ. 2008.** Species persistence in northerly glacial refugia of Europe: a matter of chance or biogeographical traits? *Journal of Biogeography* **35:** 464-482.

**Bilton DT, Mirol PM, Mascheretti S, Fredga K, Zima J, Searle JB. 1998.** Mediterranean Europe as an area of endemism for small mammals rather than a source for northwards postglacial colonization. *Proceedings of the Royal Society of London, Series B: Biological Sciences* **265:** 1219-1226.

**Böhme MU, Fritz U, Kotenko T, Džukić G, Ljubisavljević K, Tzankov N, Berendonk TU. 2007.** Phylogeography and cryptic variation within the *Lacerta viridis* complex (Lacertidae, Reptilia). *Zoologica Scripta* **36:** 119-131.

**Bonardi G, Cavazza W, Perrone V, Rossi S. 2001.** Calabria-Peloritani terrane and Northern Ionian Sea. In: Vai GB, Martini IP (Eds.). *Anatomy of an orogen: The Apennines and adjacent Mediterranean basins*. Kluwer. Dordrecht, The Netherlands. Pp. 287-306

**Bonfiglio L, Mangano G, Marra AC, Masini F, Pavia M, Petruso D. 2002.** Pleistocene Calabrian and Sicilian bioprovinces. *Geobios* **35:** 29-39.

**Bookstein FL. 1982.** Foundations of morphometrics. *Annual Review of Ecology and Systematics* **13:** 451-470.

**Bookstein FL. 1989.** Principal warps: Thin-Plate Splines and the decomposition of deformations. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **11:** 567-585.

**Bookstein FL. 1991** *Morphometric tools for landmark data. Geometry and biology*. Cambridge University Press. Cambridge, UK.

**Bookstein FL. 1996.** Combining the tools of geometric morphometrics. In: Marcus LF, Corti M, Loy A, Naylor GJP, Slice DE. (Eds). *Advances in morphometrics. NATO ASI Series*. Plenum Press. New York, USA. Pp. 131-152.

**Brito P, Edwards S. 2009.** Multilocus phylogeography and phylogenetics using sequence-based markers. *Genetica* **135:** 439-455.

**Brunhoff C, Galbreath KE, Fedorov VB, Cook JA, Jaarola M. 2003.** Holarctic phylogeography of the root vole (*Microtus oeconomus*): implications for late Quaternary biogeography of high latitudes. *Molecular Ecology* **12:** 957-968.

**Bužan E, Kryštufek B, Bryja J. In press.** Microsatellite markers confirm extensive population fragmentation of the endangered Balkan palaeoendemic Martino's vole (*Dinaromys bogdanovi*). *Conservation Genetics*.

**Caloi L, Malatesta A, Palombo RM. 1989.** Biogeografia della Calabria meridionale durante il Quaternario. *Atti Accademia Peloritana dei Pericolanti Classe I, Scienze Matematiche Fisiche e Naturali* **67:** 261-278.

**Canestrelli D, Cimmaruta R, Nascetti G. 2007.** Phylogeography and historical demography of the Italian treefrog, *Hyla intermedia*, reveals multiple refugia, population expansions and secondary contacts within peninsular Italy. *Molecular Ecology* **16:** 4808-4821.

**Canestrelli D, Cimmaruta R, Nascetti G. 2008.** Population genetic structure and diversity of the Apennine endemic stream frog, *Rana italica*; insights on the Pleistocene evolutionary history of the Italian peninsular biota. *Molecular Ecology* **17:** 3856-3872.

**Cardini A. 2003.** The geometry of the marmot (Rodentia: Sciuridae) mandible: phylogeny and patterns of morphological evolution. *Systematic Biology* **52:** 186-205.

**Cardini A, O'Higgins P. 2004.** Patterns of morphological evolution in *Marmota* (Rodentia, Sciuridae): geometric morphometrics of the cranium in the context of marmot phylogeny, ecology and conservation. *Biological Journal of the Linnean Society* **82:** 385-407.

**Cardini A, Jansson A-U, Elton S. 2007.** A geometric morphometric approach to the study of ecogeographical and clinal variation in vervet monkeys. *Journal of Biogeography* **34:** 1663-1678.

**Carraway LN, Verts BJ. 2005.** Assessment of variation in cranial and mandibular dimensions in geographic races of *Sorex trowbridgii*. In: Merritt JF, Churchfield S, Hutterer R, Sheftel BI (Eds.). *Advances in the biology of shrews II*. Special Publication of the International Society of Shrew Biologists. New York, USA. Pp. 139-154.

**Castién E, Gosálbez J. 1999.** Habitat and food preferences in a guild of insectivorous mammals in the Western Pyrenees. *Acta Theriologica* **44:** 1-13.

**Castiglia R, Annesi F, Aloise G, Amori G. 2007.** Mitochondrial DNA reveals different phylogeographic structures in the water shrews *Neomys anomalus* and *N. fodiens* (Insectivora: Soricidae) in Europe. *Journal of Zoological Systematics and Evolutionary Research* **45:** 255-262.

**Castiglia R, Annesi F, Aloise G, Amori G. 2008.** Systematics of the *Microtus savii* complex (Rodentia, Cricetidae) via mitochondrial DNA analyses: paraphyly and pattern of sex chromosome evolution. *Molecular Phylogenetics and Evolution* **46:** 1157-1164.

**Caumul R, Polly PD. 2005.** Phylogenetic and environmental components of morphological variation: skull, mandible, and molar shape in marmots (*Marmota*, Rodentia). *Evolution* **59:** 2460-2472.

**Churchfield S. 1990.** *The natural history of shrews.* Christopher Helm and A&C Black. London, UK.

**Cook LM. 1961.** The edge effect in population genetics. *The American Naturalist* **95:** 295-307.

**Corbet GB. 1961.** Origin of the British insular races of small mammals and of the 'Lusitanian' fauna. *Nature* **191:** 1037-1040.

**Corbet GB, Harris S. 1991.** *The handbook of British mammals.* Blackwell Science. Oxford, UK.

**Corti M, Aguilera M, Capanna E. 2001.** Size and shape changes in the skull accompanying speciation of South American spiny rats (Rodentia: *Proechimys* spp.). *Journal of Zoology* **253:** 537-547.

**Crowcroft P. 1957.** *The life of the shrew.* Max Reinhardt. London, UK.

**Cuenca-Bescós G, Rofes J, López-García JM, Blain H-A, De Marfá RJ, Galindo-Pellicena MA, Bennásar-Serra ML, Melero-Rubio M, Arsuaga JL,**

**Bermúdez de Castro JM, Carbonell E. 2010.** Biochronology of Spanish Quaternary small vertebrate faunas. *Quaternary International* **212:** 109-119.

**Davis MB, Shaw RG. 2001.** Range shifts and adaptive responses to Quaternary climate change. *Science* **292:** 673-679.

**Deffontaine V, Libois R, Kotlík P, Sommer R, Nieberding C, Paradis E, Searle JB, Michaux JR. 2005.** Beyond the Mediterranean peninsulas: evidence of Central European glacial refugia for a temperate forest mammal species, the bank vole (*Clethrionomys glareolus*). *Molecular Ecology* **14:** 1727-1739.

**Delany MJ, Healy MJR. 1966.** Variation in the white-toothed shrews (*Crocidura* spp.) in the British Isles. *Proceedings of the Royal Society of London, Series B: Biological Sciences* **164:** 63-74.

**Domínguez-Domínguez O, Vázquez-Domínguez E. 2009.** *Filogeografía: aplicaciones en taxonomía y conservación*. Museo de Zoología: Barcelona, Spain.

**Drummond A, Rambaut A. 2007.** BEAST: Bayesian Evolutionary Analysis by Sampling Trees. *BMC Evolutionary Biology* **7:** 214.

**Drummond AJ, Rambaut A, Shapiro B, Pybus OG. 2005.** Bayesian coalescent inference of past population dynamics from molecular sequences. *Molecular Biology and Evolution* **22:** 1185-1192.

**Drummond AJ, Ho SYW, Phillips MJ, Rambaut A. 2006.** Relaxed phylogenetics and dating with confidence. *PLoS Biology* **4:** e88.

**Dryden IL, Mardia KV. 1998.** *Statistical shape analysis*. John Wiley and Sons. Chichester, UK.

**Dubey S, Zaitsev M, Cosson J-F, Abdukadier A, Vogel P. 2006.** Pliocene and Pleistocene diversification and multiple refugia in a Eurasian shrew (*Crocidura suaveolens* group). *Molecular Phylogenetics and Evolution* **38:** 635-647.

**Dubey S, Salamin N, Ohdachi SD, Barrière P, Vogel P. 2007.** Molecular phylogenetics of shrews (Mammalia: Soricidae) reveal timing of transcontinental colonizations. *Molecular Phylogenetics and Evolution* **44:** 126-137.

**Edwards SV. 1997.** Relevance of microevolutionary processes to higher level molecular systematics. In: Mindell DP (Ed.). *Avian molecular evolution and systematics.* Academic Press. New York, USA. Pp. 251-274.

**Ehlers J. 1996.** *Quaternary and glacial geology.* John Wiley and Sons. New York, USA.

**Ehlers J. and Gibbard PL. 2004.** *Quaternary glaciations. Extent and chronology. Part 1: Europe.* Elsevier. New York, USA.

**Esposti MD, De Vries S, Crimi M, Ghelli A, Patarnello T, Meyer A. 1993.** Mitochondrial cytochrome *b*: evolution and structure of the protein. *Biochimica et Biophysica Acta (BBA) - Bioenergetics* **1143:** 243-271.

**Excoffier L, Smouse PE, Quattro JM. 1992.** Analysis of molecular variance inferred from metric distances among DNA haplotypes: Application to human mitochondrial DNA restriction data. *Genetics* **131:** 479-491.

**Excoffier L, Laval G, Schneider S. 2005.** Arlequin ver. 3.0: An integrated software package for population genetics data analysis. *Evolutionary Bioinformatics* **1:** 47-50.

**Fadda C, Corti M. 2001.** Three-dimensional geometric morphometrics of *Arvicanthis*: implications for systematics and taxonomy. *Journal of Zoological Systematics and Evolutionary Research* **39:** 235-245.

**Felsenstein J. 2004.** *Inferring phylogenies.* Sinauer Associates. Sunderland, USA.

**Fløjgaard C, Normand S, Skov F, Svenning J-C. 2009.** Ice age distributions of European small mammals: insights from species distribution modelling. *Journal of Biogeography* **36:** 1152-1163.

**Fontanillas P, Dépraz A, Giorgi MS, Perrin N. 2005.** Nonshivering thermogenesis capacity associated to mitochondrial DNA haplotypes and gender in the greater white-toothed shrew, *Crocidura russula*. *Molecular Ecology* **14:** 661-670.

**Forster P, Toth A. 2003.** Toward a phylogenetic chronology of ancient Gaulish, Celtic, and Indo-European. *Proceedings of the National Academy of Sciences of the United States of America* **100:** 9079-9084.

**Forster P, Romano V, Cali F, Röhl A, Hurles M. 2004.** MtDNA markers for Celtic and Germanic language areas in the British Isles. In: Jones M. (Ed.) *Traces of ancestry: studies in honour of Colin Renfrew.* McDonald Institute for Archaeological Research. Cambridge, UK. Pp. 99-111.

**Foulds LR, Hendy MD, Penny D. 1979.** A graph theoretic approach to the development of minimal phylogenetic trees. *Journal of Molecular Evolution* **13:** 127-149.

**Frafjord K. 2008.** Can environmental factors explain size variation in the common shrew (*Sorex araneus*)? *Mammalian Biology - Zeitschrift fur Saugetierkunde* **73:** 415-422.

**Fu YX. 1997.** Statistical tests of neutrality of mutations against population growth, hitchhiking and background selection. *Genetics* **147:** 915-925.

**Gaffney V, Thomson K, Fitch S. 2007.** *Mapping Doggerland: the Mesolithic landscapes of the southern North Sea.* Archaeopress. Oxford, UK.

**Garnier S, Magniez-Jannin F, Rasplus J-Y, Alibert P. 2005.** When morphometry meets genetics: inferring the phylogeography of *Carabus solieri* using Fourier analyses of pronotum and male genitalia. *Journal of Evolutionary Biology* **18:** 269-280.

**Gascuel O. 1997.** BIONJ: an improved version of the NJ algorithm based on a simple model of sequence data. *Molecular Biology and Evolution* **14:** 685-695.

**Gibbard PL, Head MJ, Walker MJC, The Subcomission of Quaternary Stratigraphy. 2010.** Formal ratification of the Quaternary System/Period and the Pleistocene Series/Epoch with a base at 2.58 Ma. *Journal of Quaternary Science* **25:** 96-102.

**Gómez A, Lunt DH. 2007.** Refugia within refugia: patterns of phylogeographic concordance in the Iberian Peninsula. In: Weiss S, Ferrand N. (Eds). *Phylogeography of southern European refugia*. Springer. Dordrecht, The Netherlands. Pp. 155-188.

**Grill A, Amori G, Aloise G, Lisi I, Tosi G, Wauters LA, Randi E. 2009.** Molecular phylogeography of European *Sciurus vulgaris*: refuge within refugia? *Molecular Ecology* **18:** 2687-2699.

**Gündüz I, Auffray J-C, Britton-Davidian J, Catalan J, Ganem G, Ramalhinho MG, Mathias ML, Searle JB. 2001.** Molecular studies on the colonization of the Madeiran archipelago by house mice. *Molecular Ecology* **10:** 2023-2029.

**Gündüz I, Jaarola M, Tez C, Yeniyurt C, Polly PD, Searle JB. 2007.** Multigenic and morphometric differentiation of ground squirrels (*Spermophilus*, Scuiridae, Rodentia) in Turkey, with a description of a new species. *Molecular Phylogenetics and Evolution* **43:** 916-935.

**Haffer J. 1969.** Speciation in Amazonian forest birds. *Science* **165:** 131-137.

**Hall TA. 1999.** BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symposium Series* **41:** 95-98.

**Hammer Ø, Harper DAT, Ryan PD. 2001.** PAST: Paleontological statistics software package for education and data analysis. *Palaeontologia Electronica* **4:** 9 pp.

**Hampe A, Petit RJ. 2005.** Conserving biodiversity under climate change: the rear edge matters. *Ecology Letters* **8:** 461-467.

**Hanski I. 1986.** Population dynamics of shrews on small islands accord with the equilibrium model. *Biological Journal of the Linnean Society* **28:** 23-36.

**Hasegawa M, Kishino H, Yano T-A. 1985.** Dating of the human-ape splitting by a molecular clock of mitochondrial DNA. *Journal of Molecular Evolution* **22:** 160-174.

**Hastings WK. 1970.** Monte Carlo sampling methods using Markov chains and their applications. *Biometrika* **57:** 97-109.

**Haynes S, Jaarola M, Searle JB. 2003.** Phylogeography of the common vole (*Microtus arvalis*) with particular emphasis on the colonization of the Orkney archipelago. *Molecular Ecology* **12:** 951-956.

**Heckel G, Burri R, Fink S, Desmet J-F, Excoffier L. 2005.** Genetic structure and colonization processes in European populations of the common vole, *Microtus arvalis. Evolution* **59:** 2231-2242.

**Hellborg L, Ellegren H. 2003.** Y chromosome conserved anchored tagged sequences (YCATS) for the analysis of mammalian male-specific DNA. *Molecular Ecology* **12:** 283-291.

**Hewitt GM. 1996.** Some genetic consequences of ice ages, and their role, in divergence and speciation. *Biological Journal of the Linnean Society* **58:** 247-276.

**Hewitt GM. 2000.** The genetic legacy of the Quaternary ice ages. *Nature* **405:** 907-913.

**Hewitt GM. 2004.** Genetic consequences of climatic oscillations in the Quaternary. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences* **359:** 183-195.

**Hickerson MJ, Carstens BC, Cavender-Bares J, Crandall KA, Graham CH, Johnson JB, Rissler L, Victoriano PF, Yoder AD. 2010.** Phylogeography's

past, present, and future: 10 years after Avise, 2000. *Molecular Phylogenetics and Evolution* **54:** 291-301.

**Hilbert DW, Graham A, Hopkins MS. 2007.** Glacial and interglacial refugia within a long-term rainforest refugium: The wet tropics bioregion of NE Queensland, Australia. *Palaeogeography, Palaeoclimatology, Palaeoecology* **251:** 104-118.

**Ho SYW, Larson G. 2006.** Molecular clocks: when times are a-changin'. *Trends in Genetics* **22:** 79-83.

**Ho SYW, Phillips MJ, Cooper A, Drummond AJ. 2005.** Time dependency of molecular rate estimates and systematic overestimation of recent divergence times. *Molecular Biology and Evolution* **22:** 1561-1568.

**Ho SYW, Saarma U, Barnett R, Haile J, Shapiro B. 2008.** The effect of inappropriate calibration: three case studies in molecular ecology. *PLoS Biology* **3:** e1615.

**Huelsenbeck JP, Ronquist F. 2001.** MrBayes: Bayesian inference of phylogenetic trees. *Bioinformatics* **17:** 754-755.

**Huelsenbeck JP, Ronquist F, Nielsen R, Bollback JP. 2001.** Bayesian inference of phylogeny and its impact on evolutionary biology. *Science* **294:** 2310-2314.

**Hutterer R. 1990.** *Sorex minutus* Linnaeus, 1766. In: Niethammer J, Krapp F. *Handbuch der Säugetiere Europas*, Vol. 3/1: Insektenfresser, Primaten. Aula Verlag. Wiesbaden, Germany. Pp. 183-206.

**Hutterer R. 2005.** Order Soricomorpha. In: Wilson DE, Reeder DM (Eds.). *Mammal species of the world. A taxonomic and geographic reference.* Johns Hopkins University Press. Baltimore, USA. Pp. 220-311.

**Hutterer R, Amori G, Kryštufek B, Fernandes M, Meinig H. 2008.** *Sorex minutus*. IUCN 2009, IUCN Red List of Threatened Species Ver. 2009.2. Available at: www.iucnredlist.org, accessed 12 May 2009.

**Jaarola M, Searle JB. 2002.** Phylogeography of field voles (*Microtus agrestis*) in Eurasia inferred from mitochondrial DNA sequences. *Molecular Ecology* **11:** 2613-2621.

**Johns GC, Avise JC. 1998.** A comparative summary of genetic distances in the vertebrates from the mitochondrial cytochrome *b* gene. *Molecular Biology and Evolution* **15:**1481-1490.

**Jones EP. 2009.** *The house mouse in the North East Atlantic region.* PhD thesis dissertation. The University of York, Department of Biology, UK.

**Jukes TH, Cantor CR. 1969.** Evolution of protein molecules. In: Munro HN (Ed.) *Mammalian protein metabolism.* Academic Press. New York, USA. Pp. 21-132.

**Kendall DG. 1977.** The diffusion of shape. *Advances in Applied Probability* **9:** 428-430.

**Klingenberg CP, Monteiro LR. 2005.** Distances and directions in multidimensional shape spaces: implications for morphometric applications. *Systematic Biology* **54:** 678-688.

**Klingenberg CP, Leamy LJ, Routman EJ, Cheverud JM. 2001.** Genetic architecture of mandible shape in mice: effects of quantitative trait loci analyzed by geometric morphometrics. *Genetics* **157:** 785-802.

**Knowles LL. 2009.** Statistical phylogeography. *Annual Review of Ecology, Evolution, and Systematics* **40:** 593-612.

**Knowles LL, Maddison WP. 2002.** Statistical phylogeography. *Molecular Ecology* **11:** 2623-2635.

**Kotlík P, Deffontaine V, Mascheretti S, Zima J, Michaux JR, Searle JB. 2006.** A northern glacial refugium for bank voles (*Clethrionomys glareolus*). *Proceedings of the National Academy of Sciences of the United States of America* **103:** 14860-14864.

**Kryštufek B, Reed JM. 2004.** Pattern and process in Balkan biodiversity - an overview. In: Griffiths HI, Kryštufek B, Reed JM (Eds.). *Balkan biodiversity. Pattern and process in the European hotspot*. Kluwer Academic Publishers. Dordrecht, The Netherlands. Pp. 1-8.

**Kryštufek B, Bužan EV, Hutchinson WF, Hänfling B. 2007.** Phylogeography of the rare Balkan endemic Martino's vole, *Dinaromys bogdanovi*, reveals strong differentiation within the Western Balkan Peninsula. *Molecular Ecology* **16:** 1221-1232.

**Kuhner MK. 2006.** LAMARC 2.0: maximum likelihood and Bayesian estimation of population parameters. *Bioinformatics* **22:** 768-770.

**Lambeck K, Purcell AP. 2001.** Sea-level change in the Irish Sea since the Last Glacial Maximum: constraints from isostatic modelling. *Journal of Quaternary Science* **16:** 497-506.

**Lanave C, Preparata G, Sacone C, Serio G. 1984.** A new method for calculating evolutionary substitution rates. *Journal of Molecular Evolution* **20:** 86-93.

**Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, Valentin F, Wallace IM, Wilm A, Lopez R, Thompson JD, Gibson TJ, Higgins DG. 2007.** Clustal W and Clustal X version 2.0. *Bioinformatics* **23:** 2947-2948.

**Lebarbenchon C, Poitevin F, Arnal V, Montgelard C. In press.** Phylogeography of the weasel (*Mustela nivalis*) in the western-Palaearctic region: combined effects of glacial events and human movements. *Heredity*.

**Librado P, Rozas J. 2009.** DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* **25:** 1451-1452.

**Lomolino MV. 2005.** Body size evolution in insular vertebrates: generality of the island rule. *Journal of Biogeography* **32:** 1683-1699.

**Loy A. 1996.** An introduction to geometric morphometrics and intraspecific variation: a fascinating adventure. In: Marcus LF, Corti M, Loy A, Naylor GJP, Slice DE (Eds.). *Advances in morphometrics.* Plenum Press. New York, USA. Pp. 271-273.

**Lynch M, Crease T. 1990.** The analysis of population survey data on DNA sequence variation. *Molecular Biology and Evolution* **7:** 377-394.

**MacDonald D, Barrett P. 1993.** *Mammals of Britain and Europe.* Harper Collins Publishers. New York, USA.

**Madureira ML, Magalhães CMP. 1980.** *Small mammals of Portugal.* Museo Bocage Arq. 7 (13) (2.s). Lisboa, Portugal.

**Magri D, Vendramin GG, Comps B, Dupanloup I, Geburek T, Gömöry D, Latałowa M, Litt T, Paule L, Roure JM, Tantau I, van der Knaap WO, Petit RJ, de Beaulieu J-L. 2006.** A new scenario for the Quaternary history of European beech populations: palaeobotanical evidence and genetic consequences. *New Phytologist* **171:** 199-221.

**Malatesta A. 1985.** *Geologia e paleobiologia dell'era glaciale*. La Nuova Italia Scientifica, Roma, Italy.

**Maridet O, Escarguel G, Costeur L, Mein P, Hugueney M, Legendre S. 2007.** Small mammal (rodents and lagomorphs) European biogeography from the Late Oligocene to the mid Pliocene. *Global Ecology and Biogeography* **16:** 529-544.

**Martínková N, McDonald RA, Searle JB. 2007.** Stoats (*Mustela erminea*) provide evidence of natural overland colonization of Ireland. *Proceedings of the Royal Society B: Biological Sciences* **274:** 1387-1393.

**Mascheretti S, Rogatcheva MB, Gündüz İ, Fredga K, Searle JB. 2003.** How did pygmy shrews colonize Ireland? Clues from a phylogenetic analysis of mitochondrial cytochrome *b* sequences. *Proceedings of the Royal Society B: Biological Sciences* **270:** 1593-1599.

**Masini F, Sala B. 2007.** Large- and small-mammal distribution patterns and chronostratigraphic boundaries from the Late Pliocene to the Middle Pleistocene of the Italian peninsula. *Quaternary International* **160:** 43-56.

**Masini F, Giannini T, Abbazzi L, Fanfani F, Delfino M, Christian Maul L, Torre D. 2005.** A latest Biharian small vertebrate fauna from the lacustrine succession of San Lorenzo (Sant'Arcangelo Basin, Basilicata, Italy). *Quaternary International* **131:** 79-93.

**Mayr E. 1963.** *Animal species and evolution*. Harvard University Press. Cambridge, USA.

**McDevitt AD, Rambau RV, O'Brien J, McDevitt CD, Hayden TJ, Searle JB. 2009.** Genetic variation in Irish pygmy shrews *Sorex minutus* (Soricomorpha: Soricidae): implications for colonization history. *Biological Journal of the Linnean Society* **97:** 918-927.

**McDevitt AD, Yannic G, Rambau RV, Hayden TJ, Searle JB. 2010.** Postglacial recolonization of continental Europe by the pygmy shrew (*Sorex minutus*) inferred from mitochondrial and Y chromosomal DNA sequences. In: Habel JC, Assmann T (Eds.). *Relict species: phylogeography and conservation biology*. Springer-Verlag. Berlin, Germany. Pp. 217-236.

**McDonald JH, Kreitman M. 1991.** Adaptive protein evolution at the Adh locus in *Drosophila*. *Nature* **351:** 652-654.

**McNeely J A, Miller KR, Reid WV, Mittermeier RA, Werner TB. 1990.** *Conserving the world's biological diversity.* World Conservation Union, World Resources Institute, Conservation International, World Wildlife Fund–US, and the World Bank. Washington, DC, USA.

**Metropolis N, Rosenbluth AW, Rosenbluth MN, Teller AH, Teller E. 1953.** Equation of state calculations by fast computing machines. *The Journal of Chemical Physics* **21:** 1087-1092.

**Michaux JR, Magnanou E, Paradis E, Nieberding C, Libois R. 2003.**
Mitochondrial phylogeography of the woodmouse (*Apodemus sylvaticus*) in the
Western Palearctic region. *Molecular Ecology* **12:** 685-697.

**Michaux J, Cucchi T, Renaud S, Garcia-Talavera F, Hutterer R. 2007.** Evolution
of an invasive rodent on an archipelago as revealed by molar shape analysis: the
house mouse in the Canary Islands. *Journal of Biogeography* **34:** 1412-1425.

**Millien V. 2006.** Morphological evolution is accelerated among island mammals.
*PLoS Biology* **4:** e321.

**Mitchell-Jones AJ, Amori G, Bogdanowicz W, Kryštufek B, Reijnders PJH,
Spitzenberger F, Stubble M, Thissen JBM, Vohralik V, Zima J. 1999.** *The
atlas of European mammals*. Academic Press. London, UK.

**Mitteroecker P, Gunz P. 2009.** Advances in geometric morphometrics.
*Evolutionary Biology* **36:** 235-247.

**Monteiro LR, dos Reis SF. 2005.** Morphological evolution in the mandible of spiny
rats, genus *Trinomys* (Rodentia: Echimyidae). *Journal of Zoological Systematics
and Evolutionary Research* **43:** 332-338.

**Monteiro LR, Duarte LC, dos Reis SF. 2003.** Environmental correlates of
geographical variation in skull and mandible shape of the punaré rat *Thrichomys
apereoides* (Rodentia: Echimyidae). *Journal of Zoology* **261:** 47-57.

**Moritz C. 1994.** Applications of mitochondrial DNA analysis in conservation: a
critical review. *Molecular Ecology Notes* **3:** 401-411.

**Myers N, Mittermeier RA, Mittermeier CG, da Fonseca GAB, Kent J. 2000.**
Biodiversity hotspots for conservation priorities. *Nature* **403:** 853-858.

**Nagorsen DW, Cardini A. 2009.** Tempo and mode of evolutionary divergence in
modern and Holocene Vancouver Island marmots (*Marmota vancouverensis*)
(Mammalia, Rodentia). *Journal of Zoological Systematics and Evolutionary
Research* **47:** 258-267.

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

**Nei M, Li WH. 1979.** Mathematical model for studying genetic variation in terms of restriction endonucleases. *Proceedings of the National Academy of Sciences of the United States of America* **76:** 5269-5273.

**Nielsen R. 2005.** Molecular signatures of natural selection. *Annual Review of Genetics* **39:** 197-218.

**Nogueira MR, Peracchi AL, Monteiro LR. 2009.** Morphological correlates of bite force and diet in the skull and mandible of phyllostomid bats. *Functional Ecology* **23:** 715-723.

**Ochocińska D, Taylor JRE. 2003.** Bergmann's rule in shrews: geographical variation of body size in Palearctic *Sorex* species. *Biological Journal of the Linnean Society* **78:** 365-381.

**Ohdachi SD, Hasegawa M, Iwasa MA, Vogel P, Oshida T, Lin L-K, Abe H. 2006.** Molecular phylogenetics of soricid shrews (Mammalia) based on mitochondrial cytochrome *b* gene sequences: with special reference to the Soricinae. *Journal of Zoology* **270:** 199-200.

**Oppenheimer S. 2006.** *The origins of the British: a genetic detective story.* Constable and Robinson. London, UK.

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

**Pergams ORW, Ashley MV. 2001.** Microevolution in island rodents. *Genetica* **112-113:** 245-256.

**Petit RJ, Aguinagalde I, de Beaulieu J-L, Bittkau C, Brewer S, Cheddadi R, Ennos R, Fineschi S, Grivet D, Lascoux M, Mohanty A, Muller-Starck G, Demesure-Musch B, Palme A, Martin JP, Rendell S, Vendramin GG. 2003.**

Glacial refugia: hotspots but not melting pots of genetic diversity. *Science* **300:** 1563-1565.

**Phillips T. 2004.** Seascapes and landscapes in Orkney and Northern Scotland. *World Archaeology* **35:** 371-384.

**Pierpaoli M, Riga F, Trocchi V, Randi E. 1999.** Species distinction and evolutionary relationships of the Italian hare (*Lepus corsicanus*) as described by mitochondrial DNA sequencing. *Molecular Ecology* **8:** 1805-1817.

**Polly PD. 2001.** On morphological clocks and paleophylogeography: towards a timescale for *Sorex* hybrid zones. *Genetica* **112-113:** 339-357.

**Polly PD. 2007.** Phylogeographic differentiation in *Sorex araneus*: morphology in relation to geography and karyotype. *Russian Journal of Theriology* **6:** 73-84.

**Posada D. 2008.** jModelTest: phylogenetic model averaging. *Molecular Biology and Evolution* **25:** 1253-1256.

**Posada D, Crandall KA. 2001.** Selecting the best-fit model of nucleotide substitution. *Systematic Biology* **50:** 580-601.

**Provan J, Bennett KD. 2008.** Phylogeographic insights into cryptic glacial refugia. *Trends in Ecology and Evolution* **23:** 564-571.

**Prugnolle F, de Meeus T. 2002.** Inferring sex-biased dispersal from population genetic tools: a review. *Heredity* **88:** 161-165.

**Ramos-Onsins SE, Rozas J. 2002.** Statistical properties of new neutrality tests against population growth. *Molecular Biology and Evolution* **19:** 2092-2100.

**Ratkiewicz M, Fedyk S, Banaszek A, Gielly L, Chetnicki W, Jadwiszczak K, Taberlet P. 2002.** The evolutionary history of the two karyotypic groups of the common shrew, *Sorex araneus*, in Poland. *Heredity* **88:** 235-242.

**Renaud S, Michaux JR. 2007.** Mandibles and molars of the wood mouse, *Apodemus sylvaticus* (L.): integrated latitudinal pattern and mosaic insular evolution. *Journal of Biogeography* **34:** 339-355.

**Reumer JWF. 1998.** A classification of the fossil and recent shrews. In: Wójcik JM, Wolsan M. (Eds.). *Evolution of shrews.* Mammal Research Institute, Polish Academy of Sciences. Białowieża, Poland. Pp. 5-21.

**Rogers AR. 1995.** Genetic evidence for a Pleistocene population explosion. *Evolution* **49:** 608-615.

**Rogers AR, Harpending H. 1992.** Population growth makes waves in the distribution of pairwise genetic differences. *Molecular Biology and Evolution* **9:** 552-569.

**Rohlf FJ. 1990.** Morphometrics. *Annual Review of Ecology and Systematics* **21:** 299-316.

**Rohlf FJ. 1993.** Relative warp analysis and an example of its application to mosquito wings. In: Marcus LF, Bello E, Garcia-Valdecasas A (Eds.). *Contributions to morphometrics.* Museo Nacional de Ciencias Naturales (CSIC), Vol. 8. Madrid, Spain. Pp. 131-159.

**Rohlf FJ, Marcus LF. 1993.** A revolution in morphometrics. *Trends in Ecology and Evolution* **8:**129-132.

**Rohlf FJ, Slice D. 1990.** Extensions of the Procrustes method for the optimal superimposition of landmarks. *Systematic Biology* **39:** 40-59.

**Rzebik-Kowalska B. 1998.** Fossil history of shrews in Europe. In: Wójcik J, Wolsan M. (Eds.). *Evolution of shrews.* Mammal Research Institute, Polish Academy of Sciences. Białowieża, Poland. Pp. 23-71.

**Rzebik-Kowalska B. 2005.** Paleontological relationships of European *Sorex.* In: Merritt JF, Churchfield S, Hutterer R, Sheftel BI (Eds.). *Advances in the Biology of Shrews II*, Special Publication of the International Society of Shrew Biologists. New York, USA. Pp. 1-8.

**Rzebik-Kowalska B. 2008.** Insectivores (Soricomorpha, Mammalia) from the Pliocene and Pleistocene of Transbaikalia and Irkutsk region (Russia). *Quaternary International* **179:** 96-100. Pp. 23-92.

**Schneider S, Excoffier L. 1999.** Estimation of past demographic parameters from the distribution of pairwise differences when the mutation rates vary among sites: application to human mitochondrial DNA. *Genetics* **152:** 1079-1089.

**Searle JB, Kotlík P, Rambau RV, Marková S, Herman JS, McDevitt AD. 2009.** The Celtic fringe of Britain: insights from small mammal phylogeography. *Proceedings of the Royal Society B: Biological Sciences* **276:** 4287-4294.

**Slatkin M. 1995.** A measure of population subdivision based on microsatellite allele frequencies. *Genetics* **139:** 457-462.

**Slatkin M, Hudson RR. 1991.** Pairwise comparisons of mitochondrial DNA sequences in stable and exponentially growing populations. *Genetics* **129:** 555-562.

**Sokal RR, Rohlf FJ. 1995.** *Biometry: the principles and practice of statistics in biological research.* 3rd. edition. WH Freeman and Co. New York, USA.

**Sommer RS, Nadachowski A. 2006.** Glacial refugia of mammals in Europe: evidence from fossil records. *Mammal Review* **36:** 251-265.

**Sparti A, Genoud M. 1989.** Basal rate metabolism and temperature regulation in *Sorex coronatus* and *S. minutus* (Soricidae: Mammalia). *Comparative Biochemistry and Physiology* **92A:** 359-363.

**Stamatakis A. 2006.** RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* **22:** 2688-2690.

**Stanhope MJ, Waddell VG, Madsen O, de Jong W, Hedges SB, Cleven GC, Kao D, Springer MS. 1998.** Molecular evidence for multiple origins of Insectivora and

for a new order of endemic African insectivore mammals. *Proceedings of the National Academy of Sciences of the United States of America* **95:** 9967-9972.

**Steinfartz S, Veith M, Tautz D. 2000.** Mitochondrial sequence analysis of *Salamandra* taxa suggests old splits of major lineages and postglacial recolonizations of Central Europe from distinct source populations of *Salamandra salamandra*. *Molecular Ecology* **9:** 397-410.

**Stewart JR, Dalén L. 2008.** Is the glacial refugium concept relevant for northern species? A comment on Pruett and Winker 2005. *Climatic Change* **86:** 19-22.

**Stewart JR, Lister AM. 2001.** Cryptic northern refugia and the origins of the modern biota. *Trends in Ecology and Evolution* **16:** 608-613.

**Stewart JR, Lister AM, Barnes I, Dalén L. 2010.** Refugia revisited: individualistic responses of species in space and time. *Proceedings of the Royal Society B: Biological Sciences* **277:** 661-671.

**Storch G, Qiu Zhuding, Zazhigin V. 1998.** Fossil history of shrews in Europe. In: Wójcik J, Wolsan M. (Eds.). *Evolution of shrews*. Mammal Research Institute, Polish Academy of Sciences. Białowieża, Poland. Pp. 93-120.

**Strait G. 1993.** Molar morphology and food texture among small-bodied insectivorous mammals. *Journal of Mammalogy* 74: 391-401.

**Stuart AJ. 1995.** Insularity and Quaternary vertebrate faunas in Britain and Ireland. *Geological Society, London, Special Publications* **96:** 111-125.

**Svenning J-C, Normand S, Kageyama M. 2008.** Glacial refugia of temperate trees in Europe: insights from species distribution modelling. *Journal of Ecology* **96:** 1117-1127.

**Swofford DL. 2000.** PAUP*. Phylogenetic Analysis Using Parsimony (*and other methods). Version 4. Sinauer Associates. Sunderland, USA.

**Taberlet P, Fumagalli L, Wust-Saucy A-G, Cosson J-F. 1998.** Comparative phylogeography and postglacial colonization routes in Europe. *Molecular Ecology* **7:** 453-464.

**Tajima F. 1989.** Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* **123:** 585-595.

**Taylor, JRE. 1998.** Evolution of energetic strategies in shrews. In: Wójcik J, Wolsan M. (Eds.). *Evolution of shrews*. Mammal Research Institute, Polish Academy of Sciences. Białowieża, Poland. Pp. 309-346.

**Temple HJ, Terry A (Eds.). 2007.** *The status and distribution of European mammals.* Office for Official Publications of the European Communities. Luxembourg, Luxembourg.

**Ungaro A, Cecchetti S, Aloise G, Nascetti G. 2001.** *Paleogeographic events in Southern Italy and the genetic structure of* Talpa romana. 8th Congress European Society Evolutionary Biology, Aarhus, Denmark 20-26 August 2001. Abstracts: 88.

**Van Valen L. 1973.** Pattern and the balance of nature. *Evolutionary Theory,* 1: 31–49.

**Vega R, Amori G, Aloise G, Cellini S, Loy A, Searle JB. 2010a.** Genetic and morphological variation in a Mediterranean glacial refugium: evidence from Italian pygmy shrews (*Sorex minutus*, Mammalia, Soricomorpha). *Biological Journal of the Linnean Society* **100:** 774-787.

**Vega R, Fløjgaard C, Lira-Noriega A, Nakazawa Y, Svenning J-C, Searle JB. 2010b.** Northern glacial refugia for the pygmy shrew *Sorex minutus* in Europe revealed by phylogeographic analyses and species distribution modelling. *Ecography* **33:** 260-271.

**Waddell PJ, Steel MA. 1997.** General Time-Reversible distances with unequal rates across sites: Mixing $\Gamma$ and inverse Gaussian distributions with invariant sites. *Molecular Phylogenetics and Evolution* **8:** 398-414.

**Weir BS, Cockerham CC. 1984.** Estimating F-statistics for the analysis of population structure. *Evolution* **38:** 1358-1370.

**Weninger B, Schulting R, Bradtmöller M, Clare L, Collard M, Edinborough K, Hilpert J, Jöris O, Niekus M, Rohling EJ, Wagner B. 2008.** The catastrophic final flooding of Doggerland by the Storegga slide tsunami. *Documenta Praehistorica* **XXXV:** 1-24.

**White TA. 2007.** *The genetics, ecology and evolution of island populations of the common shrew,* Sorex araneus. PhD thesis dissertation. The University of York, Department of Biology, UK.

**White TA, Searle JB. 2007.** Genetic diversity and population size: island populations of the common shrew, *Sorex araneus. Molecular Ecology* **16:** 2005-2016.

**White TA, Searle JB. 2008a.** The colonization of Scottish islands by the common shrew, *Sorex araneus* (Eulipotyphla: Soricidae). *Biological Journal of the Linnean Society* **94:** 797-808.

**White TA, Searle JB. 2008b.** Mandible asymmetry and genetic diversity in island populations of the common shrew, *Sorex araneus. Journal of Evolutionary Biology* **21:** 636-641.

**White TA, Searle JB. 2009.** Ecomorphometric variation and sexual dimorphism in the common shrew (*Sorex araneus*). *Journal of Evolutionary Biology* **22:** 1163-1171.

**Willis KJ, van Andel TH. 2004.** Trees or no trees? The environments of Central and Eastern Europe during the last glaciation. *Quaternary Science Reviews* **23:** 2369-2387.

**Willis KJ, Rudner E, Sümegi P. 2000.** The full-glacial forests of Central and South-Eastern Europe. *Quaternary Research* **53:** 203-213.

**Wilson AC, Cann RL, Carr SM, George M, Gyllensten UB, Helm-Bychowski KM, Higuchi RG, Palumbi SR, Prager EM, Sage RD, Stoneking M. 1985.** Mitochondrial DNA and two perspectives on evolutionary genetics. *Biological Journal of the Linnean Society* **26:** 375-400.

**Wójcik JM, Kawako A, Marková S, Searle JB, Kotlík P. 2010.** Phylogeographic signatures of northward post-glacial colonization from high-latitude refugia: a case study of bank voles using museum specimens. *Journal of Zoology* **281:** 249-262.

**Wright S. 1969.** *Evolution and the genetics of populations, Vol. II. The theory of gene frequencies.* University of Chicago Press. Chicago, USA.

**Xia X. 1998.** The rate heterogeneity of nonsynonymous substitutions in mammalian mitochondrial genes. *Molecular Biology and Evolution* **15:** 336-344.

**Yalden DW. 1982.** When did the mammal fauna of the British Isles arrive? *Mammal Review* **12:** 1-56.

**Yalden DW. 2007.** Zoological perspectives on the late glacial. In: Pettitt P, Bahn P, Ripoll S (Eds.). *Palaeolithic cave art at Creswell Crags in European context.* Oxford University Press. Oxford, UK. Pp. 53-60.

**Yannic G, Basset P, Hausser J. 2008.** A new perspective on the evolutionary history of western European *Sorex araneus* group revealed by paternal and maternal molecular markers. *Molecular Phylogenetics and Evolution* **47:** 237-250.

**Zelditch ML, Swiderski DL, Sheets HD, Fink WL. 2004.** *Geometric morphometrics for biologists: a primer.* Elsevier Academic Press. London, UK.

**Zvelebil M. 2006.** Mobility, contact, and exchange in the Baltic Sea basin 6000-2000 BC. *Journal of Anthropological Archaeology* **25:** 178-192.

# Genetic and morphological variation in a Mediterranean glacial refugium: evidence from Italian pygmy shrews, *Sorex minutus* (Mammalia: Soricomorpha)

RODRIGO VEGA[1,2], GIOVANNI AMORI[3], GAETANO ALOISE[4], SIMONETTA CELLINI[3], ANNA LOY[5] and JEREMY B. SEARLE[1]*

[1]*Department of Biology, University of York, PO Box 373, York YO10 5YW, UK*
[2]*Department of Entomology, Comstock Hall 5123, Cornell University, Ithaca, NY 14853, USA*
[3]*CNR, Institute of Ecosystem Studies, Via A. Borelli 50, 00161 Rome, Italy*
[4]*Dipartimento di Ecologia, Università della Calabria, Via P. Bucci, s.n., 87036 Rende, Italy*
[5]*Dipartimento di Scienze e Tecnologie per l'Ambiente e il Territorio, Università del Molise, Via Mazzini 8, I-86170 Isernia, Italy*

At the Last Glacial Maximum (LGM), the southern European peninsulas were important refugia for temperate species. Current genetic subdivision of species within these peninsulas may reflect past population subdivision at the LGM, as in 'refugia within refugia', and/or at other time periods. In the present study, we assess whether pygmy shrew populations from different regions within Italy are genetically and morphologically distinct. One maternally and two paternally inherited molecular markers (cytochrome *b* and Y-chromosome introns, respectively) were analysed using several phylogenetic methods. A geometric morphometric analysis was performed on mandibles to evaluate size and shape variability between populations. Mandible shape was also explored with a functional approach that considered the mandible as a first-order lever affecting bite force. We found genetically and morphologically distinct European, Italian, and southern Italian groups. Mandible size increased with decreasing latitude and southern Italian pygmy shrews exhibited mandibles with the strongest bite force. It is not clear whether or not the southern Italian and Italian groups of pygmy shrews occupied different refugia within the Italian peninsula at the LGM. It is likely, however, that geographic isolation earlier than the LGM on islands at the site of present-day Calabria was important in generating the distinctive southern Italian group of pygmy shrews, and also the genetic groups in other small vertebrates that we review here. Calabria is an important hotspot for genetic diversity, and is worthy of conservation attention.   © 2010 The Linnean Society of London, *Biological Journal of the Linnean Society*, 2010, **100**, 774–787.

ADDITIONAL KEYWORDS: Calabria – cytochrome *b* – geometric morphometrics – refugia – Italy – phylogeography – Y-chromosome introns.

## INTRODUCTION

The three southern European peninsulas, namely the Iberian, Italian, and Balkan (Fig. 1), acted as refugial areas for many species during the Quaternary ice ages, and are consequently species-rich areas and current hotspots of intra-specific diversity (Bilton *et al*., 1998; Hewitt, 2000; Myers *et al*., 2000; Petit *et al*., 2003). Traditionally, the southern peninsulas have been considered as single refugial areas at the Last Glacial Maximum (LGM; Bilton *et al*., 1998; Petit *et al*., 2003). However, recent studies indicate that species in the southern European peninsulas may have persisted in multiple distinct and separated

*Corresponding author. E-mail: jbs3@york.ac.uk
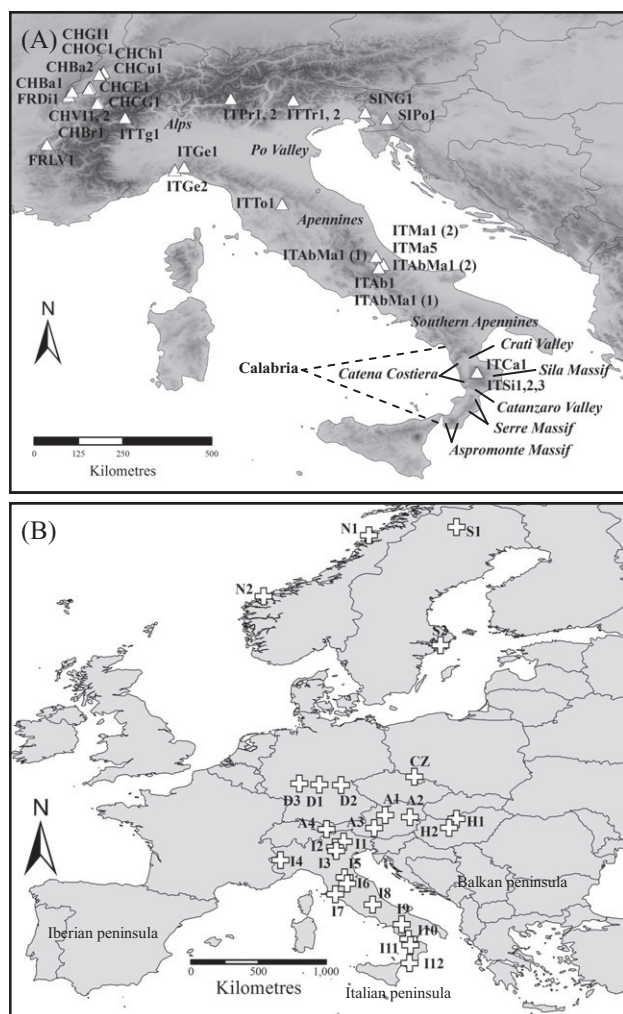
**Figure 1.** Sample localities for the pygmy shrew in Italy and elsewhere for (A) the cytochrome *b* data, with relief illustrated in grey tones, and (B) the morphometric data (samples from western Siberia are not mapped).

glacial refugia in Iberia (Gómez & Lunt, 2006), Italy (Canestrelli, Cimmaruta & Nascetti, 2007, 2008), and the Balkans (Kryštufek *et al.*, 2007). If this pattern is common, it has important consequences for the interpretation of European phylogeography as well as our understanding of biological diversity.

In the present study, we consider genetic and morphological subdivision in one species of small mammal in the Italian refugial area and compare this with other small vertebrate species from the same area. Our focal species is the pygmy shrew, *Sorex minutus* Linnaeus 1766 (Mammalia, Soricomorpha), which has a wide Palaearctic distribution extending between north-western Spain and Lake Baikal in Siberia (Hutterer, 2005). The distribution of the species becomes patchy and limited to higher altitudes in southern Europe and taxonomists have associated this

trend with differentiation. Hutterer (1990) recognized five valid subspecies: *Sorex minutus minutus* (northern and central Europe to Siberia), *Sorex minutus gymnurus* Chaworth–Musters, 1932 (Greece), *Sorex minutus becki* von Lehmann, 1963 (the Alps), *Sorex minutus carpetanus* Rey, 1971 (Spain), and *Sorex minutus lucanius* Miller, 1909 (the Basilicata and Calabria ridges of southern Italy; Fig. 1A). However, he did not resolve the taxonomic status of the pygmy shrew populations in central Italy.

Phylogeographic analyses of the pygmy shrew using mitochondrial DNA (mtDNA) and Y-chromosome introns have revealed a genetic structure over Eurasia that is considered to reflect isolation and differentiation in different refugia (Bilton *et al.*, 1998; Mascheretti *et al.*, 2003; McDevitt *et al.*, 2010). At least one glacial refugium in Italy has been proposed to explain the genetic distinctiveness of pygmy shrews existing there compared to the rest of Eurasia, and McDevitt *et al.* (2010) described a single individual with distinct mtDNA and Y-chromosome introns from southern Italy.

Thus, studies on the pygmy shrew provide evidence of lineage diversification of Italian populations from the rest of Eurasia and likely genetic and morphological subdivision within Italy. To extend these findings, we used two approaches: First, as previously employed, we implemented a phylogeographic approach, which is useful for exploring the principles and processes that generated the geographic distribution of genealogical lineages (Avise, 2000). The phylogenetic relationships among pygmy shrew populations were determined by an analysis of maternally and paternally inherited markers, adding new samples to the pre-existing data. Additionally, we used a geometric morphometrics approach (Rohlf & Marcus, 1993) on pygmy shrew mandibles from throughout Italy and neighbouring countries. Geometric morphometrics is one of the most powerful techniques for the description and interpretation of patterns of variation below the species level (Loy, 1996; Zelditch *et al.*, 2004). Because populations of pygmy shrews are scattered across a variety of different environments in Italy, adaptation may be a major driving force of recent morphological evolution in this region. Therefore, morphological variability was studied in the context of a functional hypothesis that envisages mandible shape as a first-order lever affecting bite force.

In the present study, we have also considered the genetic and morphological variation in pygmy shrews from Italy in the wider context of European biodiversity. The Italian peninsula hosts a large number of terrestrial species of mammals (74 native species), five of which are endemic, thus accentuating Italy's role and responsibility in European conservation (Gippoliti & Amori, 2002). Our phylogeographic and

morphological study, in conjunction with other data on small mammals that we review here, has important implications for the conservation of Italian mammals.

## MATERIAL AND METHODS

### PHYLOGEOGRAPHIC ANALYSIS

In total, 35 pygmy shrews were used for the cytochrome *b* (cyt *b*) analysis (Fig. 1A; see also Supporting Information, Table S1). Tissue samples were obtained from 16 individuals from several regions in northern, central, and southern Italy, plus 14 individuals from neighbouring parts of France, Switzerland, and Slovenia. Five published cyt *b* sequences of *S. minutus* (AJ535420–AJ535424) from Italy, France, and Switzerland and a sequence of *Sorex volnuchini* (used as an outgroup; AJ535458) were obtained from GenBank (Mascheretti *et al.*, 2003). Additionally, two Y-chromosome intron sequences (DBY-7 and UTY-11; Hellborg & Ellegren, 2003) were obtained from 15 male *S. minutus* from Italy and neighbouring regions, plus five DBY-7 and UTY-11 *S. minutus* sequences and a sequence of *S. volnuchini* (used as an outgroup) from McDevitt *et al.* (2010) (see Supporting Information, Table S1).

Genomic DNA was extracted using a commercial kit (Qiagen). Partial cyt *b* sequences were obtained by polymerase chain reaction (PCR) amplification either using two primer pairs that amplified approximately 700 bp of overlapping fragments or using five primer pairs (for museum samples with highly degraded DNA) that amplified approximately 250 bp of overlapping fragments (see Supporting Information, Table S2). PCR amplification for cyt *b* was performed in a 50-μL final volume: 1× buffer, 1 μM each primer, 1 μM dNTPs, 3 mM MgCl$_2$ and 0.5 U Platinum *Taq* Polymerase (Invitrogen), with cycling conditions of 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 conducted with a commercial kit (Qiagen) and sequenced (Macrogen and Cornell University Core Laboratories Center). Amplification conditions of the Y-chromosome introns are described elsewhere (Hellborg & Ellegren, 2003) and the sequences were concatenated. Known female and male samples were used as controls in all PCR reactions for Y-chromosome introns.

Sequences were edited in BIOEDIT, version 7.0.9.0 (Hall, 1999), aligned in CLUSTALX, version 2.0 (Larkin *et al.*, 2007), and collapsed into haplotypes using DNASP, version 4.90.1 (Rozas *et al.*, 2003). The model of evolution that best fit the molecular data was determined using MODELTEST, version 3.7 (Posada & Crandall, 1998), using the minimum Akaike's information criterion value. For cyt *b*, the supported substitution model was Hasegawa–Kishino–Yano 1985 with a proportion of invariable sites of 0.5308, a gamma correction of 0.8018, and nucleotide frequencies of $A = 0.2699$, $C = 0.2981$, $G = 0.1417$, and $T = 0.2903$. For the concatenated Y-chromosome introns, the best model was Tamura–Nei 1993 with an equal proportion of invariable sites, no gamma correction, and nucleotide frequencies of $A = 0.3313$, $C = 0.1957$, $G = 0.1901$, and $T = 0.2828$. The phylogenetic relationships among cyt *b* and Y-chromosome intron haplotypes were inferred using different methods in PAUP*, version 4.0b10 (Swofford, 2000), complemented with PAUPUp, version 1.0.3.1 (Calendini & Martin, 2005): neighbour-joining (NJ) and maximum likelihood (ML), using the appropriate evolutionary model, and maximum parsimony (MP), using a heuristic search with simple stepwise addition of taxa and branch swapping (tree bisection reconnection, one million rearrangements). Statistical support for the phylogenetic relationships was assessed by 10 000 bootstrap replicates for NJ and 1000 bootstrap replicates for MP and ML. A Bayesian analysis was performed in MrBayes, version 3.1 (Huelsenbeck & Ronquist, 2001), using the appropriate evolutionary model. Two runs were performed with one million generations, a sampling frequency every 100 generations (to give a total of 10 000 samples for each run), a temperature of 0.1 for the heated chain, and checked for convergence. Trees were summarized after a burn-in value of 2500 to obtain the posterior probabilities of each phylogenetic split. A phylogenetic network was constructed using NETWORK, version 4.5.1.0 (Fluxus-Engineering), with a median-joining algorithm and a greedy genetic distance calculation method.

The cyt *b* haplotypes that clustered into major genetic lineages with high bootstrap support (or high posterior probabilities) were considered as distinct phylogroups (Avise, 2000). Genetic diversity values (π, nucleotide diversity of Jukes–Cantor) for each phylogroup and for the total sample were calculated using DNASP. Genetic divergence values between all pairs of phylogroups were estimated as *Da* (the mean ± SD number of net nucleotide substitutions per site between phylogroups with Jukes–Cantor correction) using DNASP. Divergence times between cyt *b* phylogroups were estimated as $T = Da/2\mu$, where $2\mu$ is the divergence rate (again given along with the SD). We used the divergence rate of 2% per Myr assuming equal rates of mtDNA sequence divergence among phylogroups (Taberlet *et al.*, 1998). ARLEQUIN, version 3.11 (Excoffier, Laval & Schneider, 2005), was used to estimate the pairwise genetic differentiation values ($F_{ST}$) between all pairs of phylogroups, for an analysis of molecular variance (AMOVA) among and

within phylogroups, and for a locus-by-locus (per nucleotide site) AMOVA. Ten thousand nonparametric permutations were performed to generate a random distribution to test the significance of the pairwise $F_{ST}$ values and covariance components of the AMOVA, and $\alpha = 0.01$ was set as the threshold for statistical significance. No estimates of genetic diversity or differentiation were made for the Y-chromosome introns because of the low number of males per clade.

## GEOMETRIC MORPHOMETRIC ANALYSIS

We examined a total of 277 mandibles of pygmy shrews from 99 different localities in Europe and western Siberia. Because most of the localities were represented by a small number of specimens, samples were pooled on the basis of geographic proximity resulting in 27 operational taxonomic units (OTUs) (Fig. 1B; see also Supporting Information, Table S3). Images of mandibles were collected using a Pixera Professional camera (Pixera Corporation) at a 1.2 million pixel resolution equipped with a Nikkor 210 mm APO MACRO lens, at a fixed distance of 93 cm. Morphological analyses were carried out using the 'tps-Series' software (developed by F.J. Rohlf, Department of Ecology and Evolution, State University of New York at Stony Brook, NY, USA; all software are available at: http://life.bio.sunysb.edu/morph/).

Fourteen landmarks were digitized from the internal side of each mandible (Fig. 2A) using tpsDig. The size of each mandible was estimated using the centroid size (CS) (i.e. the square root of the sum of squared distances between each landmark and the centroid) (Bookstein, 1991), obtained by the software tpsRelw and was natural log-transformed. The landmark configurations were scaled, translated, and rotated using generalized Procrustes analysis (GPA; Rohlf & Slice, 1990). A weight matrix ($W$) incorporating uniform ($N = 2$) and non-uniform ($N = 22$) components was extracted using GPA (Bookstein, 1996). Both components were interpreted as shape variables ($N = 24$) and then reduced through a principal component analysis [namely, relative warp (RW) analysis] using tpsRelw.

Because of the relatively small number of individuals in each OTU, we first tested for size and shape differences between the sexes to justify pooling of sexes in subsequent analyses. We analyzed the effect of sex, OTUs, and their mutual interaction in the OTUs CZ and I9, which had an adequate number of males ($N = 13$ and $N = 12$, respectively) and females ($N = 7$ and $N = 15$, respectively). The effect of sexual dimorphism was examined by analysis of variance (ANOVA) on CS (for size) and by multivariate analysis of variance (MANOVA) on $W$ (for shape).
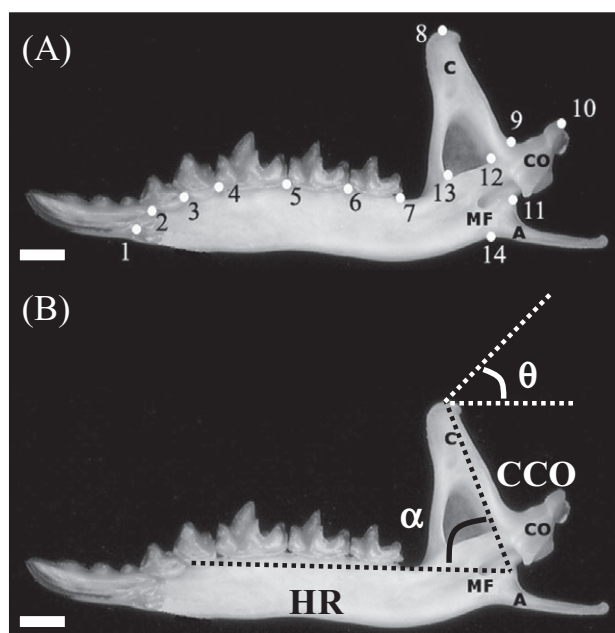


**Figure 2.** Location of 14 landmarks placed on the internal side of mandibles of the pygmy shrew (A) and measurements for bite force as a first-order lever on pygmy shrew mandibles (B), where coronoid–condyloid length (CCO) is the in-lever and horizontal ramus length (HR) is the out-lever, $\alpha$ is the angle between CCO and HR, and $\theta = 90° - \alpha$. The bite force is measured as $(\cos\theta \times CCO)/HR$. MF, mandibular fossa; C, coronoid process; CO, condyloid process; A, angular process; white bar = 1 mm.

The pattern of shape variation related to mandible size change (allometry) was compared for the full data set using tpsRegr. A correlation was performed to assess how OTUs mean CS varied with latitude (in decimals). The size differences among OTUs and among phylogroups (dividing the full mandible data set into phylogroups detected by the phylogeographic approach) were evaluated by ANOVAs and visualized with box plots. Levene's tests were performed to detect heteroscedasticity; however, the ANOVA is sufficiently robust to the violation of the assumption of homogeneity of variances (Zelditch *et al.*, 2004). We performed a Tukey–Kramer post-hoc test because it allows for unequal sample size (Sokal & Rohlf, 1995). Shape differences among OTUs and phylogroups were evaluated via MANOVAs on $W$, followed by a Hotelling $T^2$ test for multivariate comparisons.

Ordination of the OTUs was obtained through RW analysis (Rohlf, 1993) on the consensus (average) configurations of the 27 OTUs using the software tpsRelw. Shape changes in the RW space were visualized as thin-plate spline deformation grids (Bookstein, 1989). Procrustes distances among the consensus configurations of the OTUs were then

computed using tpsSmall and entered into NTSYS, version 2.2 (Exeter Software), to produce a dendrogram using an unweighted pair group method with arithmetic mean (UPGMA) to evaluate the phenetic relationships.

A functional adaptive hypothesis for shape variation in the mandible (represented as thin-plate spline deformation grids) was investigated by estimating the bite force (BF) in the context of a first-order lever (Fig. 2B; Fearnhead, Shute & Bellairs, 1955; MacDonald & Burns, 1975; Carraway & Verts, 1994), where the in-lever is the coronoid–condyloid length (CCO) measured from the tip of the coronoid process (C; landmark 8) to the base of the condyloid process (CO; landmark 11), and where the out-lever is the horizontal ramus length (HR) measured from CO to the facet-tip of $M_1$ (landmark 3; Fig. 2B). A stronger bite is given by the increased ratio between the in-lever (CCO) and the out-lever (HR), or by altering the angle ($\alpha$) between them such that it becomes more obtuse, as described by BF = $(\cos\theta \times$ CCO)/HR, where $\theta = 90° - \alpha$. CCO, HR, and $\alpha$ were measured on the deformation grids for the extreme of shape variation along the first RW (RW1) (i.e. the one that explained most of the variation).

## RESULTS

### PHYLOGEOGRAPHIC ANALYSIS

We analysed a cyt *b* fragment of 1110 bp for the 35 pygmy shrews. There were 31 haplotypes (27 haplotypes first reported in the present study; GenBank accession numbers: GQ272492–GQ272518) with 103 polymorphic sites (105 mutations and one complex codon, with three mutations not included in the analysis) of which 48 were parsimony informative. In total, there were 92 synonymous and ten nonsynonymous changes.

For the Y-chromosome introns, we analysed a concatenated fragment of 1143 bp. We found only five haplotypes that together had 11 polymorphic sites (of which nine were parsimony informative) and one site with an insertion/deletion (GenBank accession numbers: GQ272519–GQ272521 for DBY-7 and GQ272522–GQ272526 for UTY-11). Two of the haplotypes have been reported previously (see Supporting Information, Table S1).

The different phylogenetic methods produced topologically similar trees for the cyt *b* data with high bootstrap and posterior probability support for particular branches (Fig. 3A), classified as: (1) a 'European' phylogroup including samples from Switzerland and strictly northern regions of Italy in the Alps (Sondrio, Piemonte and Trento); (2) an 'Italian' phylogroup (northern–central Italy) including samples from central regions (Abruzzo) and north-western regions (Genova) in the Apennines, but also including samples from the north and north-east of Italy (Alps), Switzerland, and neighbouring regions in France and Slovenia; and (3) a 'southern Italian' phylogroup with samples strictly from Calabria (La Sila mountain). Of all the phylogroups, the southern Italian had the greatest bootstrap and posterior probability support with the various phylogenetic methods. The sequences from the whole Italian peninsula (including both the 'Italian' and 'southern Italian' phylogroups) also formed a well supported clade. One sample from Switzerland (CHCE1) clustered separately from all the rest and has been identified as belonging to a distinct Balkans lineage (R. Vega & J. B. Searle, unpublished data), and is not considered further here. The same phylogroups were also seen in the phylogenetic network (data not shown), with the European phylogroup separated from the Italian and southern Italian phylogroups by 18 mutational steps and the southern Italian and Italian phylogroups separated by 12 steps.

Similarly, the phylogenetic analysis of concatenated Y-chromosome introns (Fig. 3B) revealed a distinct 'southern Italian' lineage (one haplotype, four samples). This was most closely related to an 'Italian' lineage (one haplotype, six samples) composed of northern–central Italian and Swiss samples that in the cyt *b* trees also formed an Italian phylogroup. Additionally, there was a 'European' lineage composed of eight Swiss samples, six of which clustered within the cyt *b* European phylogroup, CHVI2 that clustered within the cyt *b* Italian phylogroup and CHCE1 that belonged to a cyt *b* Balkans lineage, plus ITPr2 (Prasota, northern Italy) that clustered within the European cyt *b* phylogroup and SIPo1 (Postjoma, Slovenia) that clustered within the Italian cyt *b* phylogroup.

The nucleotide diversity ($\pi$) was 0.0157 for the cyt *b* data overall. For the European phylogroup, $\pi = 0.0057$ relating to 11 haplotypes (all nonredundant sequences). For the Italian phylogroup, $\pi = 0.0063$ and there were 15 haplotypes, two occurring in multiple individuals (samples from Abruzzo). For the southern Italian phylogroup, $\pi = 0.0073$, reflecting four different haplotypes, although all samples were caught in the same area (La Sila mountain) and had 14 mutations that were not shared with other phylogroups (six of which were significantly differentiated from the rest of the sample by a locus-by-locus AMOVA and sequence comparison). For the whole Italian phylogroup (Italian plus southern Italian), $\pi = 0.0096$ relating to 19 haplotypes.

The divergence between phylogroups was small, as expected for within species comparisons. For European versus Italian $Da = 1.6\%$ ($\pm 0.3\%$), comparable to
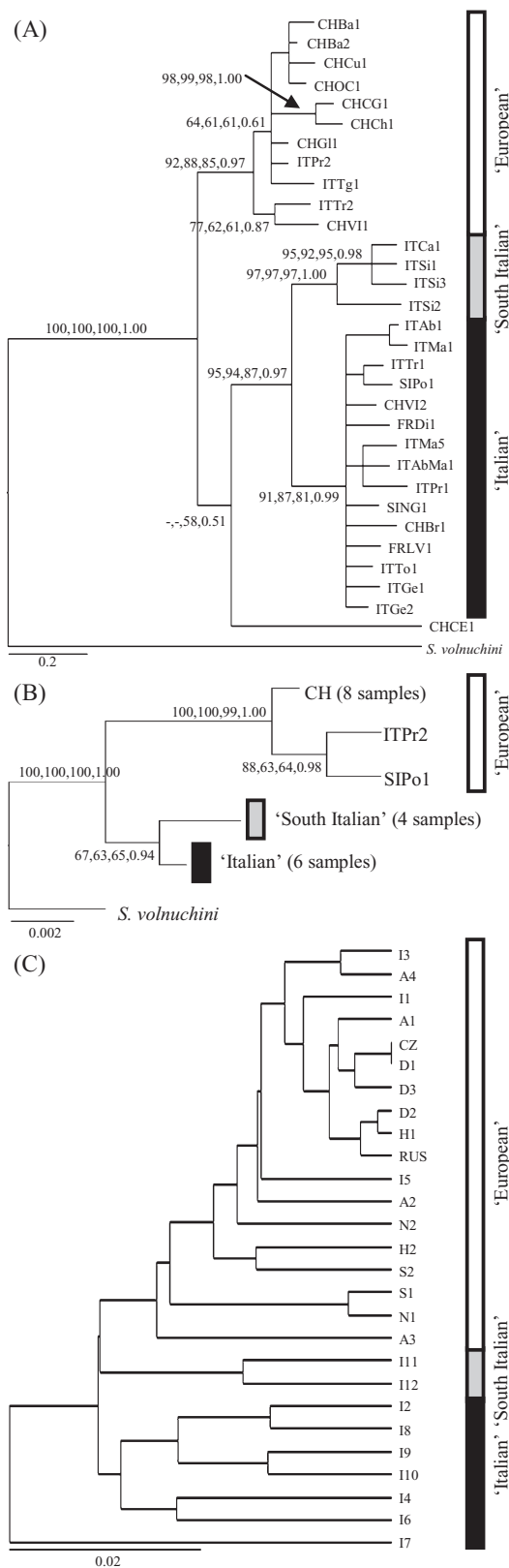
**Figure 3.** Bayesian analysis of cytochrome *b* data (A) and concatenated Y-chromosome introns DBY-7 and UTY-11 (B) for pygmy shrews with distinct lineages highlighted. C, unweighted pair group method with arithmetic mean phenogram of Proscrustes distances among the consensus configurations of 27 operational taxonomic units showing shape differences between 'European', 'Italian' and 'southern Italian' pygmy shrew mandibles. Values on phylogenetic trees correspond to branch support for neighbour-joining, maximum parsimony maximum likelihood and Bayesian analysis, respectively.

European versus southern Italian $Da = 1.5\%$ ($\pm 0.4\%$), whereas Italian versus southern Italian had the smallest divergence, $Da = 1.0\%$ ($\pm 0.3\%$). For the European versus whole Italian sample (Italian plus southern Italian) $Da = 1.4\%$ ($\pm 0.2\%$). All divergence times among phylogroups pre-dated the LGM according to the divergence rate used, where $T_{(European-Italian)} = 0.8 \pm 0.15$ Mya, $T_{(European-southern\ Italian)} = 0.75 \pm 0.2$ Mya, $T_{(Italian-southern\ Italian)} = 0.5 \pm 0.15$ Mya, and $T_{(European-whole\ Italian)} = 0.7 \pm 0.3$ Mya. The AMOVA showed that 71% of the overall genetic variation could be attributed to differences among the three phylogroups and 29% to within-phylogroup variation ($F_{ST} = 0.7101$, $P < 0.001$). Pairwise $F_{ST}$ values were high and significant between phylogroups as a result of the 27 significantly differentiated polymorphic nucleotide sites among populations as determined by the locus-by-locus AMOVA and sequence comparison.

## GEOMETRIC MORPHOMETRIC ANALYSIS

The MANOVA on the $W$ matrix revealed significant mandible shape differences between the OTUs CZ and I9 but no significant effect of sex and the interaction sex × OTU. Therefore, all subsequent analyses on shape were performed pooling all samples irrespective of sex.

The regression of shape on size (allometry) was significant ($F_{24, 6600} = 3.03$; $P < 0.001$). The ANOVA on CS among OTUs was significant ($F_{26, 273} = 12.647$, $P < 0.001$); however, Levene's test was significant and so this result has to be taken with caution and no post-hoc comparisons among OTUs were made. The ANOVA on CS among the mandible sample divided into phylogroups was significant ($F_{2, 271} = 53.212$, $P < 0.001$), and homogeneity of variances supported. Post-hoc tests showed significant CS differences among all pairs of phylogroups, although caution is warranted because of the limited number of mandibles available for the southern Italian phylogroup. There was a significant inverse correlation between mandible size and latitude ($r = -0.661$; $P < 0.001$).
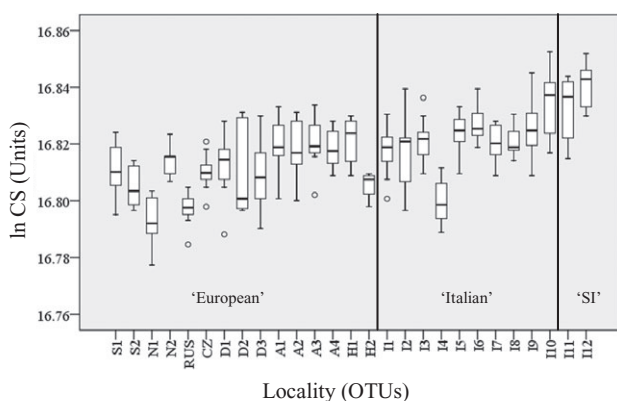
**Figure 4.** Box-plot showing mandible centroid size (CS) variation (natural log-transformed) of 27 operational taxonomic units of pygmy shrew ordered according to latitude from north (left) to south (right), and by phylogroups (SI, southern Italy).

Mandible size increased progressively from northern to southern European localities (Fig. 4). The only obvious outlier was the north-western Italian locality I4 (Torino), which fell within the range of northern–central European populations.

The MANOVA on $W$ among OTUs were significant (Wilk's $\lambda = 0.008$, $P < 0.001$). The Hotelling $T^2$ test showed that 51% of the between group comparisons were significant. The highest percentage (80–100%) of significant differences involved comparisons of central–southern Italian OTUs both among themselves and against northern–central European OTUs. The MANOVA on $W$ divided into phylogroups was significant (Wilk's $\lambda = 1.592$; $P = 0.0057$); however, Hotelling's $T^2$ test showed significant shape differences only between the European and Italian phylogroups ($P = 0.0027$). The southern Italian phylogroup was not different from either of these groups ($P = 0.1467$), most likely as a result of small sample size.

The first three RWs computed on the OTU consensuses explained cumulatively 61.75% of the total variance (28.90%, 20.51%, and 12.34%, respectively). There were two groups identifiable based on the geographic origin of the OTUs (Fig. 5A): a European group including all OTUs from northern and central Europe plus north-eastern Italian OTUs I1 and I3 (Trentino, Alps) and central Italian OTU I5 (Foreste Casentinesi, Apennines), and an Italian group with OTUs I2, I4, I6, and I8 to I12. OTU I7 (Grosseto, Tuscany) was an outlier. There was a higher diversity for the mandible within the Italian group along RW2 than within the European group. However, the southern Italian OTUs (I11 and I12) could not be distinguished from the central and northern Italian OTUs based solely on RW analysis.

The main shape differences between the two groups found with RW analysis (Italian and European) are shown, as an example, by the deformation grids taken from the most extreme values along the RW1 axis (the RW that explained most of the variation) from each group (Fig. 5B, C). There was a stronger BF in Italian compared to European pygmy shrews (BF = 0.238 and 0.074, respectively). The mandible of Italian (Fig. 5B) and European shrews (Fig. 5C) was characterized by different slopes of the incisive alveolus (landmark 2), a different HR length, and by notable changes in the posterior portion affecting the relative position of the masseteric fossa (MF) and the angular process (A). The greatest BF in Italian shrews was given by an apparent lengthening of C and a shorter HR length, which increased the ratio between the in-lever and out-lever compared to European pygmy shrew mandibles. However, in Italian shrews, there were relative movements of CO (landmarks 9, 10, and 11) and C in opposite directions causing a forward shift of its tip (landmark 8) and a more acute angle α (between CCO and HR) compared to European pygmy shrew mandibles.

The UPGMA phenogram showed morphological clusters that divided the samples into three main distinct geographical regions (Fig. 3C): (1) a 'European' cluster with northern–central European and north-eastern (Alpine) Italian OTUs (I1 and I3), but which also included OTU I5 (central Italy); (2) an 'Italian' cluster with central Italian OTUs and Italian OTUs from the southern Apennines (Collemeluccio, Muro Lucano and Catena Costiera); and (3) a strictly 'southern Italian' cluster with Calabrian samples (OTUs I11 and I12); OTU I7 (Grosseto, Tuscany) was an outlier.

## DISCUSSION

### MORPHOLOGICAL AND PHYLOGEOGRAPHICAL CONGRUENCIES

The present study represents the first combined phylogeographic and morphological analysis of the pygmy shrew, and it had a focus on populations of the species in the Italian peninsula. In general, we found congruence between phylogeographic and morphological patterns. Both analyses consistently identified pygmy shrews from the Italian peninsula as being different from northern–central European pygmy shrews, and further distinguished southern Italian pygmy shrews of the Calabria region (Fig. 1A) as being different from Italian populations further north.

We found three main geographically coherent genetic and phenetic clusters (Fig. 3A, B, C). The European morphological group (based on geometric
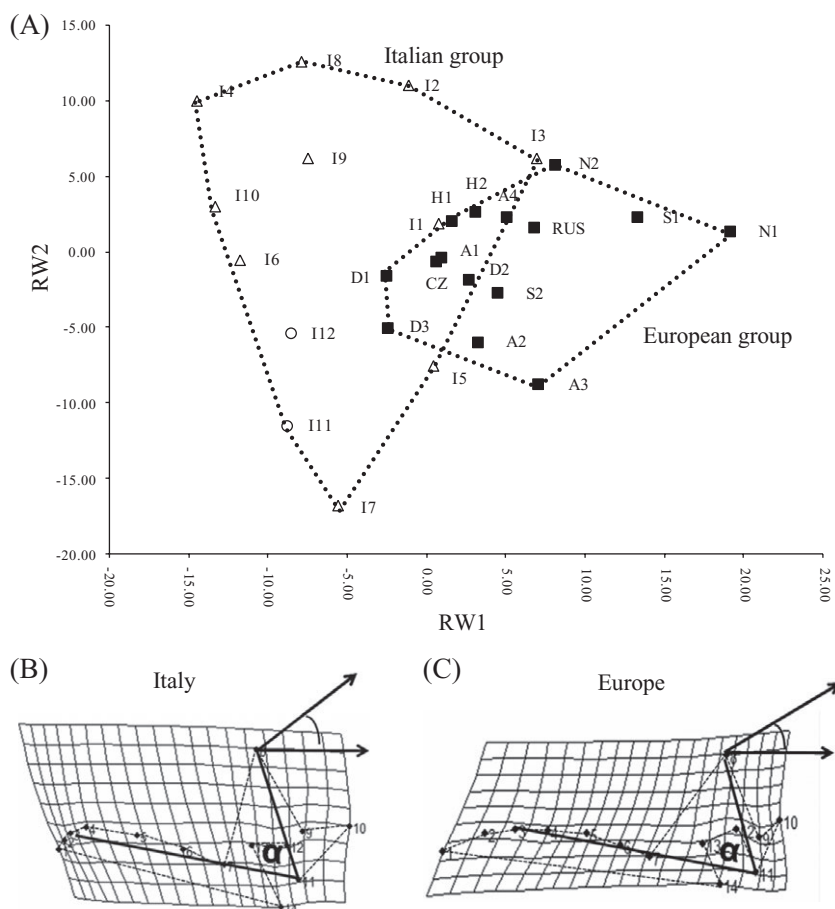
**Figure 5.** Morphometric analyses of the mandibles of pygmy shrews. Ordination of the consensus configurations (A) of 27 operational taxonomic units (OTUs) along the first two relative warp (RW) axes. ■, Northern–central European OTUs; Δ, northern–central Italian OTUs; ○, southern Italian OTUs. Shape changes and bite force implied by the variation along RW1 are shown as thin plate spline deformation grids representing the extremes of variation along the axis for (B) an Italian OTU and (C) a European OTU, respectively.

morphometrics) had a widespread distribution in Eurasia (from western Siberia to Norway and Hungary) and included samples from the northern regions of Italy. Previous data show that the European genetic group (based on cyt *b* and the Y-chromosome introns), which we found in central European and north Italian samples, is similarly very widespread (Bilton *et al*., 1998; McDevitt *et al*., 2010; Mascheretti *et al*., 2003). The Italian phenetic cluster included mandible samples from central regions of the peninsula together with OTUs from the Italian Alps and from the southern Apennines. Likewise, the Italian genetic group consisted of north Italian samples from the Alps and central Italian samples from the Apennines. However, more genetic sampling is desirable to determine whether this phylogroup reaches the southern Apennines. The situation in northern parts of Italy is also worth further study. Genetic and morphological samples from

geographically close localities in northern Italy either belonged to the European or to the Italian group (e.g. for the genetic data, samples from Trento and Prasota; for the morphological data, OTUs I1, I2, and I3). Apparently this is a contact area of two lineages.

Interestingly, we found a genetically distinct phylogroup of pygmy shrews from the Calabria region in southern Italy for cyt *b* and Y-chromosome data (all samples from La Sila). We also found a morphologically distinct group of pygmy shrews from this region, specifically from La Sila (I11) and Aspromonte (I12) massifs. When the morphological samples were pooled into the same groupings as the genetic phylogroups, the southern Italian samples had significantly larger mandibles by centroid size than other phylogroups. The southern Italian morphological samples also clustered separately by shape but, when there was no a priori grouping of OTUs, these

samples were not significantly different from northern–central Italian samples by MANOVA and RW analysis. The southern Italian morphological samples from Collemeluccio (I8), Muro Lucano (I9), and Catena Costiera (I10) clustered with northern–central Italian samples and not with those from La Sila and Aspromonte, suggesting that the morphologically and genetically distinct population of pygmy shrews in southern Italy has a very limited distribution (Fig. 1A). More extensive sampling in this southern region (Catena Costiera, other parts of the southern Apennines in Basilicata region, and further south from La Sila in the Serre and Aspromonte massifs) is desirable to obtain a more precise morphological, phylogeographic, and population genetic description. Nevertheless, the genetic results obtained in the present study indicate strongly the distinctiveness of the southern Italian phylogroup. The phylogenetic results were consistent among methods and the phylogroups displayed high branch support and were significantly differentiated. Moreover, in the few samples from southern Italy, there were several fixed and significantly differentiated mutations that were not shared with other phylogroups, and even the less variable Y-chromosome introns revealed a distinct southern Italian lineage. Collectively, the results display the phylogeographic and morphological distinctiveness of pygmy shrew populations from the Italian peninsula and the Calabria within it.

We found some discrepancy between the morphological and genetic analyses. For the morphological data, OTU I5 from central Italy (Foreste Casentinesi) clustered within the European group, whereas, for the genetic data, no central Italian sample clustered within the European lineage. The Italian cyt *b* phylogroup contained several central European samples from Switzerland, Slovenia, and France, whereas no European morphological sample grouped within the Italian cluster. Also, MANOVAs on mandibles among OTUs mainly differentiated European from central–southern Italian samples and the RW analysis discriminated these two main groups only, whereas pairwise $F_{ST}$ on genetic data showed significant differences among European, Italian, and southern Italian lineages.

OTU I7 was an outlier in the cluster and RW analyses and, to further understand this, we need better sampling along the Tyrrhenian coast. Discrepancies among maternally and paternally inherited markers, specifically in northern Italy and Switzerland, may be explained by the larger dispersal rate and activity areas of male versus female pygmy shrews (Shchipanov *et al.*, 2005), as expressed in the contact zone of distinct lineages, and by different mutation rates of the markers examined.

## SIZE AND BITE FORCE AMONG MORPHOLOGICAL GROUPS

The mandible of pygmy shrews increases in size from north to south and represents an exception to Bergmann's rule, as reported for other Soricinae in Europe (Ochocinska & Taylor, 2003), where it was suggested that small body size is adaptive under conditions of low resource availability in northern latitudes, especially in winter.

The mandible of shrews may be considered as a first-order lever during bite and mastication (MacDonald & Burns, 1975). This is because the lower condyloid facets act as the fulcrum with the in-lever (CCO) set at an acute angle to the out-lever (HR). The mandible of Italian pygmy shrews, in comparison to other European pygmy shrews, was characterized by a different slope of the incisive alveolus and a bigger ratio of the coronoid process versus the horizontal ramus lengths leading to an increased bite force, and by substantial changes in the posterior portion that created a more acute angle between the condyloid process and the horizontal ramus (Fig. 5B, C). The greater bite force of Italian pygmy shrews compared to other European pygmy shrews is likely a consequence of adaptation to the more arid conditions and prey with harder exoskeletons (Strait, 1993; Carraway & Verts, 1994). An alternative hypothesis is that inter- or intra-specific competition might cause an increase in mandible size and bite strength (Corti & Rohlf, 2001). However, considering the low population densities of pygmy shrews in southern Italy and the similarity of species assemblages throughout the animal's range, this hypothesis appears unlikely.

## REFUGIA WITHIN REFUGIA IN THE ITALIAN PENINSULA?

One possible explanation for the current occurrence of two morphologically and genetically distinct clusters of pygmy shrew essentially restricted to Italy is that there were two glacial refugia for this species in this region, and that these distinct groups arose within those refugia during one or more glacial cycles. Thus, for the pygmy shrew, the Italian refugial area may have been subdivided into multiple refugia at the LGM, concordant with the 'refugia within refugia' concept (Gómez & Lunt, 2006). However, genetic subdivision may also arise from population subdivision at times other than glacial maxima.

One of the two groups of pygmy shrews in Italy is very widespread (the Italian group), whereas the other is limited to the extreme south (the southern Italian group). It is apparent that a variety of other small vertebrates are also characterized by a genetic lineage in the extreme south of Italy (Table 1) and, in most of

**Table 1.** Small vertebrate species with a distinct genetic lineage in the southern part of the Italian peninsula

| Class | Species | Distribution range of species | Southern Italian lineage | Reference |
|---|---|---|---|---|
| Amphibia | *Triturus italicus* | Central–southern Italy | Calabria | Ragghianti & Wake (1986) |
| | *Rana* (*Pelophylax*) *lessonae* | Europe | Calabria and Sicily | Santucci *et al*. (1996); Canestrelli & Nascetti (2008) |
| | *Salamandra salamandra* | Europe | Southern Italy | Steinfartz, Veith & Tautz (2000) |
| | *Salamandrina terdigitata* | Central–southern Italy | Southern Italy | Mattoccia, Romano & Sbordoni (2005); Nascetti, Zangari & Canestrelli (2005); Canestrelli, Zangari & Nascetti (2006a) |
| | *Bombina pachypus* | Peninsular Italy | Calabria | Canestrelli *et al*. (2006b) |
| | *Hyla intermedia* | Peninsular Italy and Sicily | Calabria and Sicily | Canestrelli *et al*. (2007) |
| | *Rana italica* | Peninsular Italy | Calabria | Canestrelli *et al*. (2008) |
| Reptilia | *Hierophis viridiflavus* | Europe | Calabria and Sicily | Nagy *et al*. (2003) |
| | *Podarcis sicula* | Italy & Balkans | Calabria | Podnar, Mayer & Tvrtkovic (2005) |
| | *Vipera aspis* | Europe | Southern Italy | Ursenbacher *et al*. (2006) |
| | *Lacerta bilineata* | Europe | Calabria | Böhme *et al*. (2007) |
| Mammalia | *Lepus corsicanus* | Central–southern Italy | Southern Italy and Sicily | Pierpaoli *et al*. (1999) |
| | *Talpa romana* | Central–southern Italy | Calabria | Ungaro *et al*. (2001) |
| | *Myodes glareolus* | Eurasia | Calabria | Amori *et al*. (2008) |
| | *Microtus brachycercus* | Central–southern Italy | Central–southern Italy | Castiglia *et al*. (2008) |
| | *Sciurus vulgaris* | Eurasia | Southern Italy | Grill *et al*. (2009) |
| | *Sorex minutus* | Eurasia | Southern Italy | Present study |

these examples, there is at least one more lineage restricted to Italy (a second, Italian-restricted lineage has not been recorded in *Salamandra salamandra*, *Hierophis viridiflavus*, *Lacerta bilineata* or *Sciurus vulgaris*). So, in the pygmy shrew, as in a range of other species, this may be the result of one glacial refugium in the extreme south of Italy, and at least one other refugium further north. In *Rana* (*Pelophylax*) *lessonae*, *Vipera aspis* and *Myodes glareolus*, there is a similar situation as in the pygmy shrew, with an Italian lineage that makes contact with one or more other European lineages in the extreme north of Italy (Santucci, Nascetti & Bullini, 1996; Deffontaine *et al*., 2005; Ursenbacher *et al*., 2006; Canestrelli & Nascetti, 2008). For each of these species, the widespread Italian lineage may be presumed to derive from a glacial refugium located somewhere within the vicinity of the Apennine mountain chain.

Although the data appear consistent with the 'refugia within refugia' concept, a degree of caution is needed for the pygmy shrew, and probably for some other species as well. The concept implies that the species were restricted to particular localized areas at the LGM, which can be termed 'refugia' (Bennett & Provan, 2008; Stewart *et al*., 2010), and that their distribution expanded on amelioration of the climate. Indeed, the pygmy shrew is currently restricted to high altitude areas in Italy because the dry, hot Mediterranean conditions in the lowlands do not suit it. Therefore, in southern Europe, the distribution of the pygmy shrew may actually be more restricted during interglacials than glacials, in line with ideas of 'interglacial refugia' (Hilbert, Graham & Hopkins, 2007).

Further samples are needed to test whether the 'refugia within refugia' concept applies to the pygmy shrew in Italy (i.e. genetic analyses are needed to demonstrate population expansions from two separate refugial areas). From the current data, it is possible that the Italian and southern Italian morphological and genetic clusters occupied a single, continuous area at the LGM with the integrity of the groups retained by a hybrid zone (Barton & Hewitt, 1985).

Thus, pygmy shrews from southern Italy are genetically and morphologically distinctive, and may

be restricted to Calabria. The general question that remains is: how did the southern Italian and Italian groups of pygmy shrew become separate entities? Here, comparison with other small vertebrates can be informative. For many of the species with a southern Italian lineage, that lineage is also restricted to the peninsula of Calabria (Fig. 1A, Table 1). Calabria consists of isolated mountain massifs separated by lowland areas. From the Pliocene to the Middle Pleistocene, at times of high sea level, those massifs would have been islands in a chain that stretched between the southernmost part of the Italian peninsula and Sicily (Malatesta, 1985; Caloi, Malatesta & Palombo, 1989; Bonardi *et al*., 2001; Bonfiglio *et al*., 2002). Since the Late Pleistocene, uplift has prevented island-formation. Considering all species that have distinct Calabrian lineages, it is reasonable to suggest that the distinction of these Calabrian forms reflects island evolution. On the basis of the molecular clock used in the present study, the separation of the Italian and southern Italian cyt *b* lineages of pygmy shrew from each other and from the European lineage extend back into the Middle Pleistocene.

Over the last glacial cycle, the putative Calabrian island forms have been limited to a mainland distribution. For the pygmy shrew at least, it is likely that the Calabrian form is geographically isolated at the present time because pygmy shrews there are apparently restricted to the Calabrian massifs. The extent to which Calabrian pygmy shrews were isolated during the period extending back to the LGM is less certain, and is of course crucial in relation to the 'refugia within refugia' concept.

Calabria forms part of the 'Calabrian Arc' and includes the separate mountain massifs of Catena Costiera in the north, La Sila in central Calabria, and the Serre and Aspromonte massifs in southern Calabria (Fig. 1A; Bonardi *et al*., 2001). Catena Costiera makes contact in the north with the southern Apennines (southern Lucania, Basilicata region) but it is separated from La Sila by the Crati Valley depression in the east and from southern Calabria by the Catanzaro Plain in the south. Therefore, pygmy shrews from Catena Costiera in the northern Calabrian Arc might have been able to remain in long-term contact with the southern Apennine populations, thus avoiding differentiation. This could explain why the morphological samples from Catena Costiera (I10) clustered with northern–central Italian and not with the geographically nearby southern Italian samples from La Sila (I11) and Aspromonte (I12). For the southern Italian pygmy shrew, the La Sila and Aspromonte massifs are geographically isolated from each other and from the Apennines.

## TAXONOMIC AND LOCAL CONSERVATION IMPLICATIONS FOR SOUTHERN ITALIAN PYGMY SHREWS

The subspecies concept has been widely criticised (Wilson & Brown, 1953), although the use of trinomials can generate interest at a local scale by giving a regional- or country-based value for biodiversity, and the recognition of subspecific levels may stimulate further investigation of those populations (Mayr, 1982; Amori, Angelici & Boitani, 1999). The morphological and genetic differentiation shown in the present study does not justify that Italian and/or southern Italian pygmy shrews be considered as separate species. However, we consider that the trinomial *S. m. lucanius* should be kept as the taxonomic name to describe the southern Italian pygmy shrews, emphasizing that it is an Evolutionarily Significant Unit (ESU; Moritz, 1994). Our data also suggest that Italian pygmy shrews form a different ESU from the European *S. m. minutus*. Given that the Italian and southern Italian pygmy shrews are distinctive forms, do they need protection? Currently, the pygmy shrew overall is characterized as a Least Concern species for conservation purposes (IUCN, 2007; http://www.iucnredlist.org/details/29667/0). However, the Italian populations deserve more attention than this implies. In particular, the distinctive Calabrian populations of pygmy shrew surely need specific protection. They occur in a very small area and, on the basis of our field efforts, are very uncommon there (R. Vega & G. Aloise, pers. observ.). Given the range of species and populations that are distinctive in Calabria (Table 1), there needs to be a clear conservation effort to protect plant and animal communities in that region.

## CONCLUSIONS

On the basis of the morphological and phylogeographic data presented here, we show three distinguishable and significantly differentiated clusters of pygmy shrews corresponding to European, Italian, and southern Italian groups. The differentiation includes adaptive features, on the basis of our analysis of mandible characteristics. The two Italian groups appear to have arisen as a consequence of geographic isolation before the LGM (one group originating on islands in the location of present-day Calabria), and it is unclear whether they were located in separate glacial refugia at the LGM. Moreover, because studies of several other species have also shown genetic and morphological distinction of populations in Calabria, this region should be a focus for conservation.

## ACKNOWLEDGEMENTS

## REFERENCES

**Amori G, Angelici FM, Boitani L. 1999.** Mammals of Italy: a revised checklist of species and subspecies (Mammalia). *Senckenbergiana Biologica* **79:** 271–286.

**Amori G, Aloise G, Annesi F, Franceschini P, Colangelo P. 2008.** Filogeografia mitocondriale di Arvicola rossastra *Myodes glareolus* in Italia. *69° Congresso Unione Zoologica Italiana*, Senigallia, September 22–25. Abstracts, 18.

**Avise JC. 2000.** *Phylogeography: the history and formation of species*. Cambridge, MA: Harvard University Press.

**Barton NH, Hewitt GM. 1985.** Analysis of hybrid zones. *Annual Review of Ecology and Systematics* **16:** 113–148.

**Bennett KD, Provan J. 2008.** What do we mean by 'refugia'? *Quaternary Science Reviews* **27:** 2449–2455.

**Bilton DT, Mirol PM, Mascheretti S, Fredga K, Zima J, Searle JB. 1998.** Mediterranean Europe as an area of endemism for small mammals rather than a source for northwards postglacial colonization. *Proceedings of the Royal Society of London Series B, Biological Sciences* **265:** 1219–1226.

**Böhme MU, Fritz U, Kotenko T, Džukić G, Ljubisavljević K, Tzankov N, Berendonk TU. 2007.** Phylogeography and cryptic variation within the *Lacerta viridis* complex (Lacertidae, Reptilia). *Zoologica Scripta* **36:** 119–131.

**Bonardi G, Cavazza W, Perrone V, Rossi S. 2001.** Calabria-Peloritani terrane and northern Ionian Sea. In: Vai GB, Martini IP, eds. *Anatomy of an orogen: the Apennines and adjacent Mediterranean basins*. Dordrecht: Kluwer, 287–306.

**Bonfiglio L, Mangano G, Marra AC, Masini F, Pavia M, Petruso D. 2002.** Pleistocene Calabrian and Sicilian bioprovinces. *Geobios* **35:** 29–39.

**Bookstein FL. 1989.** Principal warps: thin-plate spline and the decomposition of transformations. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **11:** 567–595.

**Bookstein FL. 1991.** *Morphometric tools for landmark data*. Cambridge: Cambridge University Press.

**Bookstein FL. 1996.** Standard formula for the uniform shape component in landmark data. In: Marcus LF, Corti M, Loy A, Naylor GJP, Slice DE, eds. *Advances in morphometrics*. New York, NY: NATO ASI Series A, Life Sciences, **284:** 153–168.

**Calendini F, Martin JF. 2005.** *PaupUP*, Version 1.0.3.1. A graphical frontend for PAUP* DOS software. Available at: http://www.agro-montpellier.fr/sppe/Recherche/JFM/PaupUp/main.htm

**Caloi L, Malatesta A, Palombo RM. 1989.** Biogeografia della Calabria meridionale durante il Quaternario. *Atti Accademia Peloritana dei Pericolanti Classe I, Scienze Matematiche Fisiche e Naturali* **67:** 261–278.

**Canestrelli D, Cimmaruta R, Costantini V, Nascetti G. 2006b.** Genetic diversity and phylogeography of the Apennine yellow-bellied toad *Bombina pachypus*, with implications for conservation. *Molecular Ecology* **15:** 3741–3754.

**Canestrelli D, Cimmaruta R, Nascetti G. 2007.** Phylogeography and historical demography of the Italian tree frog, *Hyla intermedia*, reveals multiple refugia, population expansions and secondary contacts within peninsular Italy. *Molecular Ecology* **16:** 4808–4821.

**Canestrelli D, Cimmaruta R, Nascetti G. 2008.** Population genetic structure and diversity of the Apennine endemic stream frog, *Rana italica* – insights on the Pleistocene evolutionary history of the Italian peninsular biota. *Molecular Ecology* **17:** 3856–3872.

**Canestrelli D, Nascetti G. 2008.** Phylogeography of the pool frog *Rana* (*Pelophylax*) *lessonae* in the Italian peninsula and Sicily: multiple refugia, glacial expansions and nuclear–mitochondrial discordance. *Journal of Biogeography* **35:** 1923–1936.

**Canestrelli D, Zangari F, Nascetti G. 2006a.** Genetic evidence for two distinct species within the Italian endemic *Salamandrina terdigitata* Bonnaterre, 1789 (Amphibia: Urodela: Salamandridae). *Herpetological Journal* **16:** 221–227.

**Carraway LN, Verts BJ. 1994.** Relationship of mandibular morphology to relative bite force in some *Sorex* from western North America. In: Merritt JF, Kirkland GL, Rose RK, eds. *Advances in the biology of shrews*. Pittsburgh, PA: Special Publication of the Carnegie Museum of Natural History, 18: 201–210.

**Castiglia R, Annesi F, Aloise G, Amori G. 2008.** Systematics of the *Microtus savii* complex (Rodentia, Cricetidae) via mitochondrial DNA analysis: paraphyly and pattern of sex chromosome evolution. *Molecular Phylogenetics and Evolution* **46:** 1157–1164.

**Corti M, Rohlf FJ. 2001.** Chromosomal speciation and phenotypic evolution in the house mouse. *Biological Journal of the Linnean Society* **73:** 99–112.

**Deffontaine V, Libois R, Kotlík P, Sommer R, Nieberding C, Paradis E, Searle JB, Michaux JR. 2005.** Beyond

the Mediterranean peninsulas: evidence of central European glacial refugia for a temperate forest mammal species, the bank vole (*Clethrionomys glareolus*). *Molecular Ecology* **14:** 1727–1739.

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

**Fearnhead RW, Shute CCD, Bellairs AA. 1955.** The temporo-mandibular joint of shrews. *Proceedings of the Zoological Society of London* **125:** 795–806.

**Gippoliti S, Amori G. 2002.** Mammal diversity and taxonomy in Italy: implications for conservation. *Journal for Nature Conservation* **10:** 133–143.

**Gómez A, Lunt DH. 2006.** Refugia within refugia: patterns of phylogeographic concordance in the Iberian Peninsula. In: Weiss S, Ferrand N, eds. *Phylogeography of southern refugia: evolutionary perspectives on the origins and conservation of European biodiversity*. Dordrecht: Springer Verlag, 155–188.

**Grill A, Amori G, Aloise G, Lisi I, Tosi G, Wauters LA, Randi E. 2009.** Molecular phylogeography of European *Sciurus vulgaris*: refuge within refugia? *Molecular Ecology* **18:** 2687–2699.

**Hall TA. 1999.** BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT/XP. *Nucleic Acids Symposium Series* **41:** 95–98.

**Hellborg L, Ellegren H. 2003.** Y-chromosome conserved anchored tagged sequences (YCATS) for the analysis of mammalian male-specific DNA. *Molecular Ecology* **12:** 283–291.

**Hewitt GM. 2000.** The genetic legacy of the Quaternary ice ages. *Nature* **405:** 907–913.

**Hilbert DW, Graham A, Hopkins MS. 2007.** Glacial and interglacial refugia within a long-term rainforest refugium: the Wet Tropics Bioregion of NE Queensland, Australia. *Palaeogeography, Palaeoclimatology, Palaeoecology* **251:** 104–118.

**Huelsenbeck JP, Ronquist F. 2001.** MrBayes: Bayesian inference of phylogeny. *Bioinformatics* **17:** 754–755.

**Hutterer R. 1990.** *Sorex minutus* Linnaeus, 1766. Zwergspitzmaus. In: Niethammer J, Krapp F, eds. *Handbuch der Säugetiere Europas, band 3*. Wiesbaden: Aula-Verlag, 183–206.

**Hutterer R. 2005.** Order Soricomorpha. In: Wilson DE, Reeder DM, eds. *Mammal species of the world: a taxonomic and geographic reference*, 3rd edn, Vol. 1. Baltimore, MD: John Hopkins University Press, 220–311.

**IUCN. 2007.** Temple HJ, Terry A, comp. *The status and distribution of European mammals*. Luxembourg: The World Conservation Union, Office for Official Publications of the European Communities.

**Kryštufek B, Buzan EV, Hutchinson WF, Hänfling B. 2007.** Phylogeography of the rare Balkan endemic Martino's vole, *Dinaromys bogdanovi*, reveals strong differentiation within the western Balkan Peninsula. *Molecular Ecology* **16:** 1221–1232.

**Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, Valentin F, Wallace IM,**

**Wilm A, Lopez R, Thompson JD, Gibson TJ, Higgins DG. 2007.** Clustal W and Clustal X version 2.0. *Bioinformatics* **23:** 2947–2948.

**Loy A. 1996.** An introduction to geometric morphometrics and intraspecific variation: a fascinating adventure. In: Marcus LF, Corti M, Loy A, Naylor GJP, Slice DE, eds. *Advances in morphometrics*. New York, NY: Plenum Press, 271–273.

**McDevitt A, Yannic G, Rambau RV, Hayden TJ, Searle JB. 2010.** Postglacial re-colonization of continental Europe by the pygmy shrew (*Sorex minutus*) inferred from mitochondrial and Y chromosomal DNA sequences. In: Habel JC, Assmann T, eds. *Relict species: phylogeography and conservation biology*. Heidelberg: Springer-Verlag, 217–236.

**MacDonald SG, Burns DM. 1975.** *Physics for the life and health sciences*. Reading, PA: Addison-Wesley Publishing Company.

**Malatesta A. 1985.** *Geologia e paleobiologia dell'era glaciale*. Roma: La Nuova Italia Scientifica.

**Mascheretti S, Rogatcheva MB, Gündüz İ, Fredga K, Searle JB. 2003.** How did pygmy shrews colonize Ireland? Clues from a phylogenetic analysis of mitochondrial cytochrome *b* sequences. *Proceedings of the Royal Society of London Series B, Biological Sciences* **270:** 1593–1599.

**Mattoccia M, Romano A, Sbordoni V. 2005.** Mitochondrial DNA sequence analysis of the spectacled salamander, *Salamandrina terdigitata* (Urodela: Salamandridae), supports the existence of two distinct species. *Zootaxa* **995:** 1–19.

**Mayr E. 1982.** Of what use are subspecies? *Auk* **99:** 593–595.

**Moritz C. 1994.** Defining 'evolutionarily significant units' for conservation. *Trends in Ecology and Evolution* **9:** 373–375.

**Myers N, Mittermeier RA, Mittermeier CG, Fonseca GAB, Kent J. 2000.** Biodiversity hotspots for conservation priorities. *Nature* **403:** 853–858.

**Nagy ZT, Joger U, Guicking D, Wink M. 2003.** Phylogeography of the European whip snake *Coluber* (*Hierophis*) *viridiflavus* as inferred from nucleotide sequences of the mitochondrial cytochrome *b* gene and ISSR genomic fingerprinting. *Biota* **3:** 109–118.

**Nascetti G, Zangari F, Canestrelli D. 2005.** The spectacled salamanders, *Salamandrina terdigitata* Lacépède, 1788 and *S. perspicillata* Savi, 1821: 1) genetic differentiation and evolutionary history. *Rendiconti Lincei: Scienze Fisiche e Naturali* **16:** 159–169.

**Ochocinska D, Taylor JRE. 2003.** Bergmann's rule in shrews: geographical variation of body size in Palaearctic *Sorex* species. *Biological Journal of the Linnean Society* **78:** 365–381.

**Petit RJ, Aguinagalde I, de Beaulieu JL, Bittkau C, Brewer S, Cheddadi R, Ennos R, Fineschi S, Grivet D, Lascoux M, Mohanty A, Müller-Starck G, Demesure-Musch B, Palmé A, Martín JP, Rendell S, Vendramin GG. 2003.** Glacial refugia: hotspots but not melting pots of genetic diversity. *Science* **300:** 1563–1565.

**Pierpaoli M, Riga F, Trocchi V, Randi E. 1999.** Species distinction and evolutionary relationships of the Italian hare (*Lepus corsicanus*) as described by mitochondrial DNA sequencing. *Molecular Ecology* **8:** 1805–1817.

**Podnar M, Mayer W, Tvrtkovic N. 2005.** Phylogeography of the Italian wall lizard, *Podarcis sicula*, as revealed by mitochondrial DNA sequences. *Molecular Ecology* **14:** 575–588.

**Posada D, Crandall KA. 1998.** Modeltest: testing the model of DNA substitution. *Bioinformatics* **14:** 817–818.

**Ragghianti M, Wake DB. 1986.** Genic variation and its evolutionary implications in the Italian newt, *Triturus italicus*. *Herpetologica* **42:** 206–214.

**Rohlf FJ. 1993.** Relative warp analysis and an example of its application to mosquito wings. In: Marcus LF, Bello E, Garcia-Valdecasas A, eds. *Contributions to morphometrics*. Madrid: Museo Nacional de Ciencias Naturales, 131–159.

**Rohlf FJ, Marcus LF. 1993.** A revolution in morphometrics. *Trends in Ecology and Evolution* **8:** 129–132.

**Rohlf FJ, Slice DE. 1990.** Extensions of the Procrustes method for the optimal superimposition of landmark. *Systematic Zoology* **39:** 40–59.

**Rozas J, Sánchez-Delbarrio JC, Messeguer X, Rozas R. 2003.** DnaSP, DNA polymorphism analyses by the coalescent and other methods. *Bioinformatics* **19:** 2496–2497.

**Santucci F, Nascetti G, Bullini L. 1996.** Hybrid zones between two genetically differentiated forms of the pond frog *Rana lessonae* in southern Italy. *Journal of Evolutionary Biology* **9:** 429–450.

**Shchipanov NA, Kalinin AA, Demidova TB, Oleinichenko VY, Aleksandrov DY, Kouptzov AV. 2005.** Population ecology of red-toothed shrews, *Sorex araneus, S. caecutiens, S. minutus*, and *S. isodon*, in central Russia. In: Merrit JF, Churchfield S, Hutterer R, Sheftel BI, eds. *Advances in the biology of shrews II*. New York, NY: International Society of Shrew Biologists, 199–214.

**Sokal RR, Rohlf FJ. 1995.** *Biometry: the principles and practice of statistics in biological research*, 3rd edn. New York, NY: Freeman.

**Steinfartz S, Veith M, Tautz D. 2000.** Mitochondrial sequence analysis of *Salamandra* taxa suggests old splits of major lineages and postglacial recolonizations of central Europe from distinct source populations of *Salamandra salamandra*. *Molecular Ecology* **9:** 397–410.

**Stewart JR, Lister AM, Barnes I, Dalén L. 2010.** Refugia revisited: individualistic responses of species in space and time. *Proceedings of the Royal Society B, Biological Sciences* **277:** 661–671.

**Strait G. 1993.** Molar morphology and food texture among small-bodied insectivorous mammals. *Journal of Mammalogy* **74:** 391–402.

**Swofford DL. 2000.** *Paup*. Phylogenetic analysis using parsimony (*and other methods),* Version 4. Sunderland, MA: Sinauer.

**Taberlet P, Fumagalli L, Wust-Saucy AG, Cosson JF. 1998.** Comparative phylogeography and postglacial colonization routes in Europe. *Molecular Ecology* **7:** 453–464.

**Ungaro A, Cecchetti S, Aloise G, Nascetti G. 2001.** Paleogeographic events in southern Italy and the genetic structure of *Talpa romana*. *Eighth Congress of the European Society for Evolutionary Biology (ESEB)*. August 20–25. Abstracts, 88. Aarhus.

**Ursenbacher S, Conelli A, Golay P, Monney JC, Zuffi MAL, Thiery G, Durand T, Fumagalli L. 2006.** Phylogeography of the asp viper (*Vipera aspis*) inferred from mitochondrial DNA sequence data: evidence for multiple Mediterranean refugial areas. *Molecular Phylogenetics and Evolution* **38:** 546–552.

**Wilson EO, Brown JWL. 1953.** The subspecies concept and its taxonomic application. *Systematic Zoology* **2:** 97–111.

**Zelditch ML, Swiderski DL, Sheets HD, Fink WL. 2004.** *Geometric morphometrics for biologists: a primer*. London: Elsevier Academic Press.

## SUPPORTING INFORMATION

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

**Table S1.** Samples and localities for phylogeographic analysis of pygmy shrews.
**Table S2.** List of primers used for the amplification of cytochrome *b* (cyt *b*) and Y-chromosome introns from pygmy shrews.
**Table S3.** Samples and localities for morphological analysis of pygmy shrews.

Please note: Wiley-Blackwell is 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.
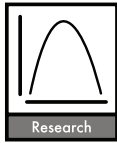
# Northern glacial refugia for the pygmy shrew *Sorex minutus* in Europe revealed by phylogeographic analyses and species distribution modelling

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

*R. Vega (rrv9@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 Rønde, 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 post-glacial (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). Further-more, 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 under-standing 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* (Mam-malia, 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 line-age persisted and expanded from one or more central or eastern European refugia located further north than the traditionally recognized southern European refugia. How-ever, 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 integra-ted 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 phylogeo-graphic 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 phylogeo-graphic 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 indepen-dent 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 wide-spread lineage present a genetic signature of population expansion? Would different SDM-based hindcasting appro-aches predict suitable LGM conditions for *S. minutus* not only in the southern European peninsulas, but also further north, consistent with northern refugia? Would the com-bined phylogeographic and SDM approach allow us to estimate more precisely the geographic locations of north-ern 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 con-tributes 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

261

(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 $MgCl_2$ 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 (Avise 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 ($\pi$, 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 Fs, a statistic based on the infinite-site model without recombination that shows large negative Fs 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 $5 \times 5$ 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 $50 \times 50$ 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 $50 \times 50$ 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 Fs ($Fs = -24.8437$,
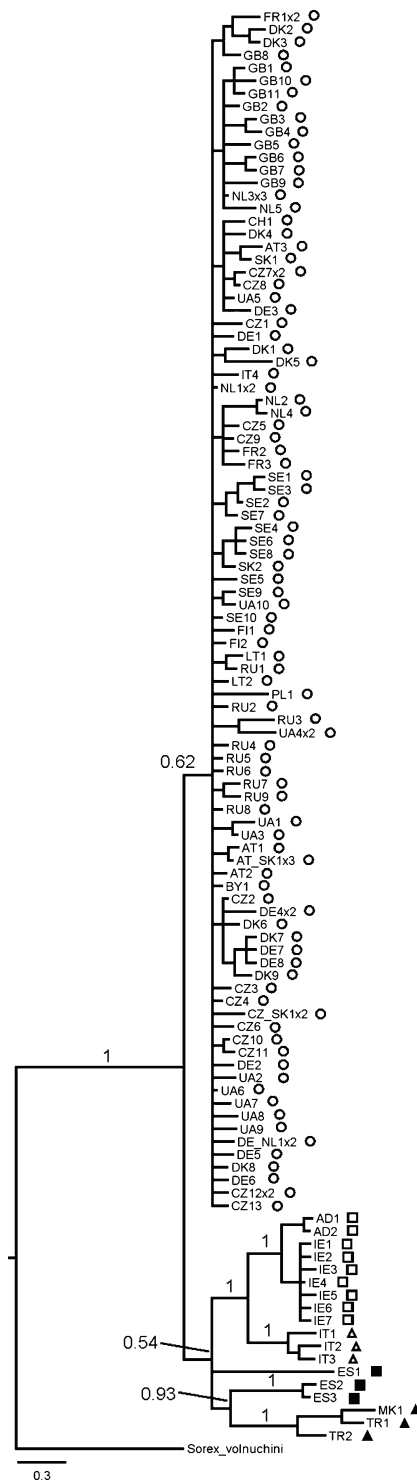
Figure 1. Continued.

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

$p < 0.001$) 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

265

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
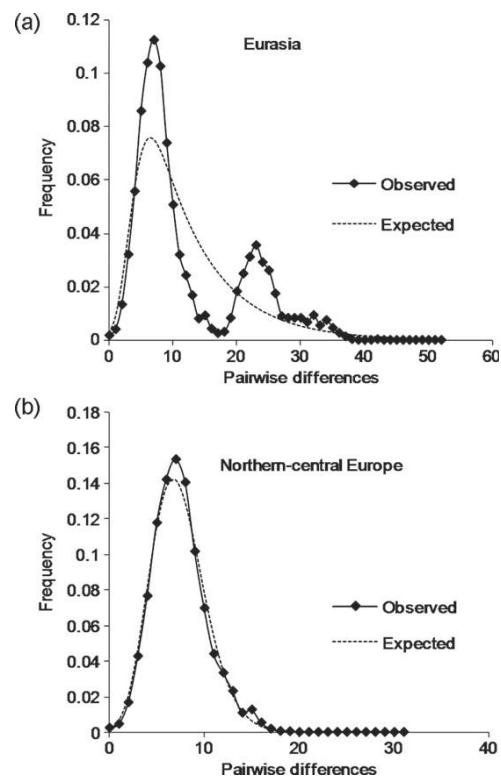


Figure 3. Mismatch distribution for observed (continuous line) and expected (dashed line) pairwise comparisons under a sudden population expansion model among *Sorex minutus* cyt b sequences. (a) Mismatch distribution among Eurasian sequences with a bimodal observed distribution where the first peak 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 (Avise 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.

266
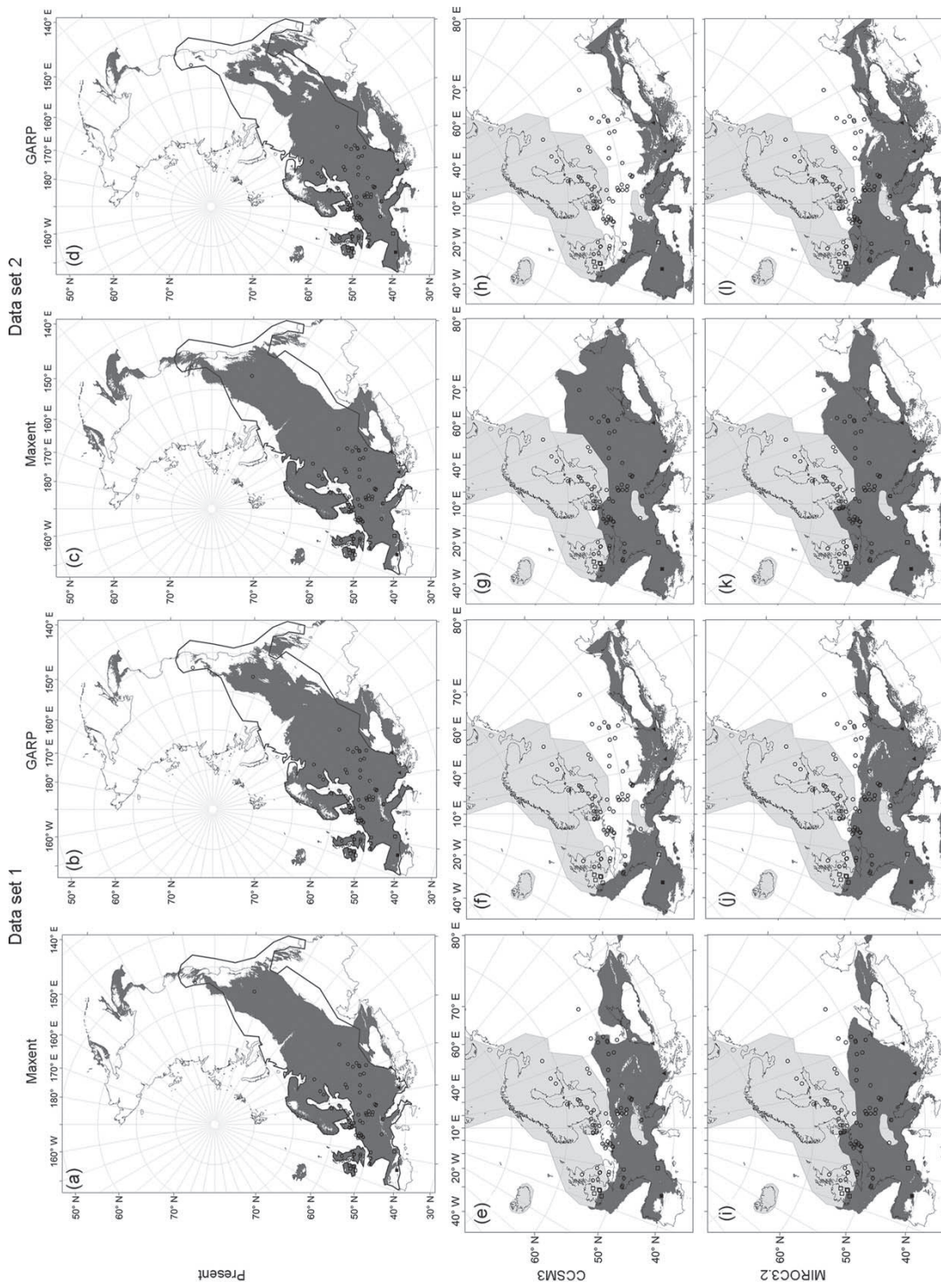
Figure 4. Continued.

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, Δ = 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.

Conference of the International Biogeography Society (8–12 January 2009, Mérida, Yucatán, México).

## 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 phylogeogaphic 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.

Samples of *Sorex minutus* for genetic analysis of cytochrome (cyt) *b*, Y-chromosome intron (Y) and BRCA1 gene.

| Country | Locality | Phylo group | Code | Longitude (Decimal) | Latitude (Decimal) | Cyt *b* hap | Y | BRCA 1 |
|---|---|---|---|---|---|---|---|---|
| Andorra | Andorra[1] | Pyr | ADAd0001 | 1.5833 | 42.5833 | 156 | | |
| Andorra | Andorra[1] | Pyr | ADAd0002 | 1.5833 | 42.5833 | 157 | | |
| Austria | Donnerskirchen[5] | NCE | ATDo1612 | 16.6413 | 47.8957 | 159 | | |
| Austria | Donnerskirchen[6] | Balk | ATDo1611 | 16.6413 | 47.8957 | 158 | 1 | 4 |
| Austria | Illmitz[5] | NCE | ATIl0003 | 16.8454 | 47.7647 | 161 | | |
| Austria | Illmitz[5] | NCE | ATIl0004 | 16.8454 | 47.7647 | 162 | 4 | |
| Austria | Illmitz[5] | NCE | ATIl0005 | 16.8454 | 47.7647 | 163 | 3 | |
| Austria | Illmitz[6] | Balk | ATIl0001 | 16.8454 | 47.7647 | 160 | 2 | |
| Austria | Illmitz[6] | Balk | ATIl0002 | 16.8454 | 47.7647 | | 3 | |
| Belorusia | Lesnojeozero[5] | NCE | BYLE0026 | 26.6918 | 54.8302 | 165 | | |
| Bosnia Herzegovina | Osjecenica[6] | Balk | BAOs5670 | 16.2887 | 44.2397 | 164 | | |
| Britain | Aird, Benbecula[3] | Pyr | GBBe0740 | -7.3328 | 57.4278 | 19 | | |
| Britain | Allt Loch an Tairbh, Rum[3] | Pyr | GBRu0105 | -6.3300 | 57.0297 | 26 | | |
| Britain | Allt nah Uamha, Rum[3] | Pyr | GBRu0104 | -6.2692 | 56.9878 | 26 | | |
| Britain | Ambergate, Derbyshire[3] | NCE | GBDe0953 | -1.4889 | 53.0619 | 37 | | |
| Britain | Balintore, Easter Ross[3] | Pyr | GBER0408 | -3.9175 | 57.7531 | 41 | | |
| Britain | Balintore, Easter Ross[3] | NCE | GBER0410 | -3.9175 | 57.7531 | 42 | | |
| Britain | Beanacre, Wiltshire[3] | NCE | GBWi0042 | -2.1494 | 51.3975 | 108 | | |
| Britain | Biggar, Walney[3] | NCE | GBWl0669 | -3.2275 | 54.0772 | 111 | | |
| Britain | Biggar, Walney[3] | NCE | GBWl0670 | -3.2275 | 54.0772 | 111 | | |
| Britain | Biggar, Walney[3] | NCE | GBWl0671 | -3.2275 | 54.0772 | 111 | | |
| Britain | Billington, Lancashire[3] | NCE | GBLa1071 | -2.4328 | 53.8122 | 70 | | |
| Britain | Birsay, OM[3] | Pyr | GBOM0261 | -3.2931 | 59.1283 | 84 | | |
| Britain | Bislington, Kent[3] | NCE | GBKe0010 | 0.9339 | 51.0672 | 66 | | |
| Britain | Bodmin, Cornwall[3] | NCE | GBCo00C1 | -4.7283 | 50.4753 | 27 | | |
| Britain | Bodmin, Cornwall[3] | Pyr | GBCo00C2 | -4.7283 | 50.4753 | 28 | | |
| Britain | Bracknell Forest[3] | NCE | GBBF0035 | -0.7436 | 51.4503 | 20 | | |
| Britain | Bracknell Forest[3] | NCE | GBBF0036 | -0.7436 | 51.4503 | 21 | | |
| Britain | Bradley Hill, Hampshire[3] | Pyr | GBHa0039 | -1.0800 | 51.1872 | 47 | | |
| Britain | Canna[3] | Pyr | GBCn0901 | -6.5117 | 57.0594 | 26 | | |
| Britain | Canna[3] | Pyr | GBCn0905 | -6.5117 | 57.0594 | 26 | | |
| Britain | Canna[3] | Pyr | GBCn0906 | -6.5117 | 57.0594 | 26 | | |
| Britain | Carinish, N. Uist[3] | Pyr | GBNU0319 | -7.3225 | 57.5181 | 19 | | |
| Britain | Carinish, N. Uist[3] | Pyr | GBNU0325 | -7.3225 | 57.5181 | 78 | | |
| Britain | Carmarthenshire[3] | NCE | GBCa0001 | -4.1553 | 51.7603 | 22 | | |
| Britain | Carmarthenshire[3] | NCE | GBCa0002 | -4.1478 | 51.7525 | 23 | | |
| Britain | Carmarthenshire[3] | NCE | GBCa0003 | -4.1478 | 51.7525 | 24 | | |
| Britain | Carmarthenshire[3] | NCE | GBCa0004 | -4.5817 | 51.7922 | 25 | | |
| Britain | Clarach, Cardiganshire[3] | NCE | GBCr1100 | -4.0611 | 52.4381 | 30 | | |
| Britain | Craighouse, Jura[3] | NCE | GBJu0815 | -5.9435 | 55.8528 | 64 | | |
| Britain | Craighouse, Jura[3] | NCE | GBJu0816 | -5.9435 | 55.8528 | 64 | | |
| Britain | Cromdale, Inverness[3] | NCE | GBIn0180 | -3.4856 | 57.3411 | 54 | | |
| Britain | Desnage, Suffolk[3] | NCE | GBSu0005 | 0.5189 | 52.2478 | 102 | | |
| Britain | Desnage, Suffolk[3] | NCE | GBSu0006 | 0.5411 | 52.2769 | 103 | | |

| Britain | Desnage, Suffolk[3] | NCE | GBSu0007 | 0.5411 | 52.2769 | 104 | | |
|---|---|---|---|---|---|---|---|---|
| Britain | Desnage, Suffolk[3] | NCE | GBSu0008 | 0.5411 | 52.2769 | 105 | | |
| Britain | Dorset[3] | NCE | GBDo0458 | -2.3633 | 50.8306 | 38 | | |
| Britain | Dorset[3] | NCE | GBDo0459 | -2.3633 | 50.8306 | 10 | | |
| Britain | Dorset[3] | NCE | GBDo0460 | -2.3633 | 50.8306 | 39 | | |
| Britain | Dorset[3] | Pyr | GBDo0461 | -2.3633 | 50.8306 | 40 | | |
| Britain | Dorset[3] | NCE | GBDo0462 | -2.3633 | 50.8306 | 38 | | |
| Britain | Dounby, OM[3] | Pyr | GBOM0419 | -3.2244 | 58.9700 | 50 | | |
| Britain | Dover, Kent[3] | NCE | GBKe0009 | 1.2578 | 51.1269 | 65 | | |
| Britain | East Cottingwith[1] | NCE | GBEC0001 | -0.9167 | 53.8667 | 233 | | |
| Britain | East Harptree, Somerset[3] | NCE | GBSo002 | -2.6597 | 51.2856 | 97 | | |
| Britain | East Hendred[1] | NCE | GBEH0001 | -1.3333 | 51.5667 | 234 | | |
| Britain | Easter Ross Balintore[3] | Pyr | GBER0177 | -3.9175 | 57.7531 | 41 | | |
| Britain | Fishnish, Mull[3] | Pyr | GBMu0858 | -5.8378 | 56.5100 | 14 | | |
| Britain | Fishnish, Mull[3] | Pyr | GBMu0859 | -5.8378 | 56.5100 | 72 | | |
| Britain | Foulney, Cumbria [3] | NCE | GBCu0642 | -3.1522 | 54.0661 | 33 | | |
| Britain | Fulford, North Yorkshire[3] | NCE | GBNY00F2 | -1.0736 | 53.9364 | 79 | | |
| Britain | Fulford, North Yorkshire[3] | NCE | GBNY00F3 | -1.0736 | 53.9364 | 80 | | |
| Britain | Fulford, North Yorkshire[3] | NCE | GBNY00Y1 | -1.0736 | 53.9364 | 81 | | |
| Britain | Gask[1] | NCE | GBGa0001 | -3.6667 | 56.3500 | 235 | | |
| Britain | Gask[1] | NCE | GBGa0002 | -3.6667 | 56.3500 | 236 | | |
| Britain | Gayton Thorpe, Norfolk[3] | NCE | GBNo0001 | 0.5919 | 52.7481 | 75 | | |
| Britain | Gayton Thorpe, Norfolk[3] | NCE | GBNo0002 | 0.5919 | 52.7481 | 10 | | |
| Britain | Gayton, Norfolk[3] | NCE | GBNo0004 | 0.5706 | 52.7628 | 77 | | |
| Britain | Gigha[3] | Pyr | GBGi0001 | -5.7410 | 55.6833 | 43 | | |
| Britain | Gigha[3] | Pyr | GBGi0002 | -5.7410 | 55.6833 | 43 | | |
| Britain | Gigha[3] | Pyr | GBGi0003 | -5.7410 | 55.6833 | 43 | | |
| Britain | Gigha[3] | Pyr | GBGi0004 | -5.7410 | 55.6833 | 43 | | |
| Britain | Gigha[3] | Pyr | GBGi0005 | -5.7410 | 55.6833 | 43 | | |
| Britain | Gigha[3] | Pyr | GBGi0006 | -5.7410 | 55.6833 | 43 | | |
| Britain | Glen Artney, Perthshire[3] | NCE | GBPr0132 | -4.0033 | 56.3464 | 90 | | |
| Britain | Gloucestershire[3] | NCE | GBGl0001 | -2.6450 | 51.7883 | 44 | | |
| Britain | Gloucestershire[3] | NCE | GBGl0043 | -2.1097 | 51.6511 | 45 | | |
| Britain | Gloucestershire[3] | NCE | GBGl0047 | -2.1031 | 51.8092 | 46 | | |
| Britain | Grimness, OS[6] | Pyr | OSGr0006 | -2.9167 | 58.8167 | 98 | | 1 |
| Britain | Grimness, OS[6] | Pyr | OSGr0007 | -2.9167 | 58.8167 | 98 | | |
| Britain | Grimness, OS[6] | Pyr | OSGr0008 | -2.9167 | 58.8167 | 98 | | |
| Britain | Grimness, OS[6] | Pyr | OSGr0009 | -2.9167 | 58.8167 | 98 | | |
| Britain | Grimness, OS[6] | Pyr | OSGr0010 | -2.9167 | 58.8167 | 98 | | |
| Britain | Grimness, OS[6] | Pyr | OSGr0011 | -2.9167 | 58.8167 | 98 | | |
| Britain | Grimness, OS[6] | Pyr | OSGr0012 | -2.9167 | 58.8167 | 98 | | |
| Britain | Grimness, OS[6] | Pyr | OSGr0013 | -2.9167 | 58.8167 | 98 | | |
| Britain | Grimness, OS[6] | Pyr | OSGr0014 | -2.9167 | 58.8167 | 98 | | |
| Britain | Grimness, OS[6] | Pyr | OSGr0015 | -2.9167 | 58.8167 | 98 | 8 | |
| Britain | Grimness, OS[6] | Pyr | OSGr0016 | -2.9167 | 58.8167 | 98 | | |
| Britain | Grimness, OS[6] | Pyr | OSGr0018 | -2.9167 | 58.8167 | 98 | 8 | |
| Britain | Grimness, OS[6] | Pyr | OSGr0020 | -2.9167 | 58.8167 | 98 | | |
| Britain | Grimness, OS[6] | Pyr | OSGr0023 | -2.9167 | 58.8167 | 98 | | |
| Britain | Grimness, OS[6] | Pyr | OSGr0025 | -2.9167 | 58.8167 | 98 | | |
| Britain | Grimness, OS[6] | Pyr | OSGr0027 | -2.9167 | 58.8167 | 98 | 8 | |
| Britain | Grimness, OS[6] | Pyr | OSGr0028 | -2.9167 | 58.8167 | 98 | | |
| Britain | Grimness, OS[6] | Pyr | OSGr0029 | -2.9167 | 58.8167 | 98 | | |
| Britain | Grimness, OS[6] | Pyr | OSGr0039 | -2.9167 | 58.8167 | 98 | | |
| Britain | Grimstone, Norfolk[3] | NCE | GBNo0003 | 0.5561 | 52.7650 | 76 | | |

| Britain | Grittenham[1] | NCE | GBGr0001 | -1.9667 | 51.5500 | 237 | | |
|---|---|---|---|---|---|---|---|---|
| Britain | Hardiwick, Staffordshire[3] | NCE | GBSt0030 | -2.0925 | 53.0983 | 101 | | |
| Britain | Harray, OM[6] | Pyr | OMHa0003 | -3.1902 | 59.0337 | 50 | | |
| Britain | Harray, OM[6] | Pyr | OMHa0004 | -3.1902 | 59.0337 | 50 | | |
| Britain | Harray, OM[6] | Pyr | OMHa0006 | -3.1902 | 59.0337 | 318 | | |
| Britain | Harray, OM[6] | Pyr | OMHa0009 | -3.1902 | 59.0337 | 318 | | |
| Britain | Harray, OM[6] | Pyr | OMHa0011 | -3.1902 | 59.0337 | 318 | | |
| Britain | Havenstreet, Isle of Wight[3] | Pyr | GBIW0001 | -1.2528 | 50.6689 | 62 | | |
| Britain | Havenstreet, Isle of Wight[3] | Pyr | GBIW0002 | -1.2528 | 50.6689 | 62 | | |
| Britain | Hestwall, OM[3] | Pyr | GBOM0001 | -2.9131 | 58.9067 | 82 | | |
| Britain | Hestwall, OM[3] | Pyr | GBOM0002 | -2.9131 | 58.9067 | 82 | | |
| Britain | Hestwall, OM[3] | Pyr | GBOM0003 | -2.9131 | 58.9067 | 82 | | |
| Britain | Hestwall, OM[3] | Pyr | GBOM0004 | -2.9131 | 58.9067 | 82 | | |
| Britain | Hobbister, OM[6] | Pyr | OMHo0014 | -3.0676 | 58.9464 | 85 | 8 | |
| Britain | Hobbister, OM[6] | Pyr | OMHo0015 | -3.0676 | 58.9464 | 85 | | |
| Britain | Holm, OM [3] | Pyr | GBOM0260 | -2.9256 | 58.9186 | 83 | | |
| Britain | Holm, OM[3] | Pyr | GBOM0262 | -2.8783 | 58.9044 | 83 | | |
| Britain | Holy Island, Anglesey[3] | NCE | GBAn0038 | -4.6667 | 53.3028 | 9 | | |
| Britain | Hoy village, Hoy[3] | Pyr | GBHo0001 | -3.3333 | 58.9167 | 50 | | |
| Britain | Islay[3] | Pyr | GBIs0001 | -6.2333 | 55.7500 | 58 | | |
| Britain | Islay[3] | Pyr | GBIs0002 | -6.2333 | 55.7500 | 59 | | |
| Britain | Islay[3] | Pyr | GBIs0003 | -6.2333 | 55.7500 | 60 | | |
| Britain | Islay[3] | Pyr | GBIs0004 | -6.2333 | 55.7500 | 61 | | |
| Britain | Isle of Man[1] | NCE | GBIM0001 | -4.4833 | 54.1500 | 238 | | |
| Britain | Isle of Man[1] | NCE | GBIM0002 | -4.4833 | 54.1500 | 239 | | |
| Britain | Isle of Man[3] | NCE | GBIM0600 | -4.4833 | 54.1500 | 52 | | |
| Britain | Isle of Man[3] | NCE | GBIM0601 | -4.3511 | 54.2914 | 53 | | |
| Britain | Isle of Man[3] | NCE | GBIM0602 | -4.4517 | 54.1725 | 52 | | |
| Britain | Kildonan, Arran [3] | Pyr | GBAr0004 | -5.1281 | 55.4500 | 16 | | |
| Britain | Kildonan, Arran [3] | Pyr | GBAr0006 | -5.1281 | 55.4500 | 17 | | |
| Britain | Kildonan, Arran [3] | Pyr | GBAr0007 | -5.1281 | 55.4500 | 18 | | |
| Britain | Kildonan, Arran [3] | Pyr | GBAr0008 | -5.1281 | 55.4500 | 17 | | |
| Britain | Kinloch, Rum[3] | Pyr | GBRu0127 | -6.2803 | 57.0125 | 93 | | |
| Britain | Kippford, Kirkcudbright[3] | NCE | GBKk0099 | -3.8233 | 54.8669 | 34 | | |
| Britain | Kippford, Kirkcudbright[3] | NCE | GBKk0100 | -3.8233 | 54.8669 | 10 | | |
| Britain | Kirbister, OM[6] | Pyr | OMKi0013 | -3.1101 | 58.9500 | 85 | | |
| Britain | Kirkcudbrightshire[3] | NCE | GBKk0589 | -3.7814 | 54.9756 | 69 | | |
| Britain | Kirkhill, Inverness[3] | NCE | GBIn0181 | -4.4169 | 57.4544 | 55 | | |
| Britain | Kirkhill, Inverness[3] | NCE | GBIn0182 | -4.4169 | 57.4544 | 56 | | |
| Britain | Knockan, Mull[3] | Pyr | GBMu0861 | -6.2006 | 56.3336 | 74 | | |
| Britain | Lang Taing, OS[3] | Pyr | GBSR0274 | -2.9214 | 58.7586 | 98 | | |
| Britain | Leicester, Leicestershire[3] | NCE | GBLe0497 | -1.1333 | 52.6333 | 71 | | |
| Britain | Llanbabo, Anglesey[3] | Pyr | GBAn0046 | -4.4353 | 53.3617 | 12 | | |
| Britain | Llangaffo, Anglesey[3] | NCE | GBAn0044 | -4.3300 | 53.1889 | 11 | | |
| Britain | Loch Swartmill, OW[6] | Pyr | OWLS0001 | -2.9334 | 59.2833 | 50 | | |
| Britain | Loch Swartmill, OW[6] | Pyr | OWLS0002 | -2.9334 | 59.2833 | 50 | | |
| Britain | Loch Swartmill, OW[6] | Pyr | OWLS0003 | -2.9334 | 59.2833 | 50 | | |
| Britain | Loch Swartmill, OW[6] | Pyr | OWLS0005 | -2.9334 | 59.2833 | 50 | | |
| Britain | Loch Swartmill, OW[6] | Pyr | OWLS0024 | -2.9334 | 59.2833 | 50 | 8 | 1 |
| Britain | Loch Swartmill, OW[6] | Pyr | OWLS0025 | -2.9334 | 59.2833 | 50 | | |

| Britain | Loch Swartmill, OW[6] | Pyr | OWLS0036 | -2.9334 | 59.2833 | 50 | | |
| Britain | Loch Swartmill, OW[6] | Pyr | OWLS0037 | -2.9334 | 59.2833 | 50 | | |
| Britain | Loch Swartmill, OW[6] | Pyr | OWLS0038 | -2.9334 | 59.2833 | 50 | | |
| Britain | Loch Swartmill, OW[6] | Pyr | OWLS0055 | -2.9334 | 59.2833 | 50 | | |
| Britain | Loch Swartmill, OW[6] | Pyr | OWLS0059 | -2.9334 | 59.2833 | 50 | | |
| Britain | Loch Swartmill, OW[6] | Pyr | OWLS0060 | -2.9334 | 59.2833 | 50 | | |
| Britain | Loch Swartmill, OW[6] | Pyr | OWLS0061 | -2.9334 | 59.2833 | 50 | | |
| Britain | Loch Swartmill, OW[6] | Pyr | OWLS0072 | -2.9334 | 59.2833 | 50 | | |
| Britain | Loch Swartmill, OW[6] | Pyr | OWLS0073 | -2.9334 | 59.2833 | 50 | | |
| Britain | Loch Swartmill, OW[6] | Pyr | OWLS0082 | -2.9334 | 59.2833 | 50 | | |
| Britain | Loch Swartmill, OW[6] | Pyr | OWLS0083 | -2.9334 | 59.2833 | 50 | | |
| Britain | Loch Swartmill, OW[6] | Pyr | OWLS0084 | -2.9334 | 59.2833 | 50 | | |
| Britain | Loch Swartmill, OW[6] | Pyr | OWLS0085 | -2.9334 | 59.2833 | 50 | | |
| Britain | Longnor[1] | NCE | GBLo0001 | -1.8833 | 53.1667 | 242 | | |
| Britain | Longtown, Cumbria[3] | NCE | GBCu0677 | -2.9686 | 55.0139 | 34 | | |
| Britain | Lyn Conwy[3] | NCE | GBCo0139 | -3.8333 | 53.2833 | 29 | | |
| Britain | Lyn Conwy[1] | NCE | GBLC0001 | -3.8333 | 53.2833 | 240 | | |
| Britain | Lyn Conwy[1] | NCE | GBLC0002 | -3.8333 | 53.2833 | 241 | | |
| Britain | Macclesfield[1] | NCE | GBMa0001 | -2.0333 | 53.2500 | 243 | | |
| Britain | Machrie, Arran[3] | Pyr | GBAr0001 | -5.2856 | 55.5564 | 13 | | |
| Britain | Machrie, Arran[3] | Pyr | GBAr0002 | -5.2856 | 55.5564 | 14 | | |
| Britain | Malltraeth, Anglesey[3] | NCE | GBAn0042 | -4.3822 | 53.1928 | 10 | | |
| Britain | Maltraeth, Anglesey[3] | NCE | GBAn0001 | -4.3403 | 53.1992 | 8 | | |
| Britain | Mull[3] | Pyr | GBMu0860 | -5.8611 | 56.4047 | 73 | | |
| Britain | Ness, OW[6] | Pyr | OWNe0002 | -2.8667 | 59.2334 | 50 | | |
| Britain | Ness, OW[6] | Pyr | OWNe0003 | -2.8667 | 59.2334 | 50 | | |
| Britain | Ness, OW[6] | Pyr | OWNe0004 | -2.8667 | 59.2334 | 50 | | |
| Britain | Ness, OW[6] | Pyr | OWNe0005 | -2.8667 | 59.2334 | 50 | | |
| Britain | Ness, OW[6] | Pyr | OWNe0006 | -2.8667 | 59.2334 | 50 | | |
| Britain | Ness, OW[6] | Pyr | OWNe0007 | -2.8667 | 59.2334 | 50 | | |
| Britain | Ness, OW[6] | Pyr | OWNe0008 | -2.8667 | 59.2334 | 50 | | |
| Britain | Ness, OW[6] | Pyr | OWNe0009 | -2.8667 | 59.2334 | 50 | | |
| Britain | Ness, OW[6] | Pyr | OWNe0026 | -2.8667 | 59.2334 | 50 | | 1 |
| Britain | Ness, OW[6] | Pyr | OWNe0028 | -2.8667 | 59.2334 | 50 | | |
| Britain | Ness, OW[6] | Pyr | OWNe0030 | -2.8667 | 59.2334 | 50 | | |
| Britain | Ness, OW[6] | Pyr | OWNe0041 | -2.8667 | 59.2334 | 50 | | |
| Britain | Ness, OW[6] | Pyr | OWNe0062 | -2.8667 | 59.2334 | 50 | 8 | |
| Britain | Ness, OW[6] | Pyr | OWNe0103 | -2.8667 | 59.2334 | 50 | | |
| Britain | Orielton, Pembrokeshire[3] | NCE | GBPe0026 | -4.9539 | 51.6508 | 88 | | |
| Britain | Orielton, Pembrokeshire[3] | NCE | GBPe0042 | -4.9539 | 51.6508 | 89 | | |
| Britain | Orphir, OM[3] | Pyr | GBOM0418 | -3.0514 | 58.9444 | 85 | | |
| Britain | Pembrokeshire[3] | NCE | GBPe0005 | -4.6575 | 51.7725 | 86 | | |
| Britain | Pembrokeshire[3] | NCE | GBPe0006 | -4.6575 | 51.7725 | 87 | | |
| Britain | Plastow, Hampshire[3] | NCE | GBHa0040 | -1.2258 | 51.3589 | 48 | | |
| Britain | Plastow, Hampshire[3] | NCE | GBHa0041 | -1.2258 | 51.3589 | 49 | | |
| Britain | Raasay[3] | Pyr | GBRa0001 | -6.0333 | 57.4000 | 91 | | |
| Britain | Raasay[3] | Pyr | GBRa0002 | -6.0333 | 57.4000 | 92 | | |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Britain | Raasay[3] | Pyr | GBRa0003 | -6.0333 | 57.4000 | 91 | | |
| Britain | Raasay[3] | Pyr | GBRa0004 | -6.0333 | 57.4000 | 91 | | |
| Britain | Rackwick Road, Hoy[3] | Pyr | GBHo0868 | -3.3828 | 58.8803 | 51 | | |
| Britain | Rinigar, OS[3] | Pyr | GBSR0612 | -3.0217 | 58.8208 | 98 | | |
| Britain | Rinigar, OS[3] | Pyr | GBSR0613 | -3.0217 | 58.8208 | 99 | | |
| Britain | Romney Warren, Kent[3] | NCE | GBKe0972 | 0.9603 | 50.9953 | 67 | | |
| Britain | Rudge, Shropshire[3] | NCE | GBSh0006 | -2.2597 | 52.5742 | 95 | | |
| Britain | Rudge, Shropshire[3] | NCE | GBSh0007 | -2.2597 | 52.5742 | 96 | | |
| Britain | Saint Ola, OM[6] | Pyr | OMSO0006 | -2.9500 | 58.9500 | 85 | | 1 |
| Britain | Saint Ola, OM[6] | Pyr | OMSO0007 | -2.9500 | 58.9500 | 83 | | |
| Britain | Saint Ola, OM[6] | Pyr | OMSO0008 | -2.9500 | 58.9500 | 83 | | 1 |
| Britain | Saint Ola, OM[6] | Pyr | OMSO0009 | -2.9500 | 58.9500 | 83 | | |
| Britain | Saint Ola, OM[6] | Pyr | OMSO0010 | -2.9500 | 58.9500 | 83 | | |
| Britain | Saint Ola, OM[6] | Pyr | OMSO0011 | -2.9500 | 58.9500 | 319 | | |
| Britain | Saint Ola, OM[6] | Pyr | OMSO0013 | -2.9500 | 58.9500 | 83 | | |
| Britain | Saint Ola, OM[6] | Pyr | OMSO0014 | -2.9500 | 58.9500 | 83 | | |
| Britain | Saint Ola, OM[6] | Pyr | OMSO0015 | -2.9500 | 58.9500 | 83 | | |
| Britain | Saint Ola, OM[6] | Pyr | OMSO0017 | -2.9500 | 58.9500 | 319 | 8 | |
| Britain | Saint Ola, OM[6] | Pyr | OMSO0020 | -2.9500 | 58.9500 | 277 | | |
| Britain | Saint Ola, OM[6] | Pyr | OMSO0026 | -2.9500 | 58.9500 | 83 | 8 | |
| Britain | Saint Ola, OM[6] | Pyr | OMSO0029 | -2.9500 | 58.9500 | 83 | | |
| Britain | Saint Ola, OM[6] | Pyr | OMSO0036 | -2.9500 | 58.9500 | 83 | | |
| Britain | Sanda Island, Kintyre[3] | Pyr | GBKi0795 | -5.5828 | 55.2819 | 68 | | |
| Britain | Sandwick, OM[6] | Pyr | OMSa0001 | -3.2972 | 59.0483 | 82 | 8 | |
| Britain | Sandwick, OM[6] | Pyr | OMSa0002 | -3.2972 | 59.0483 | 50 | 8 | |
| Britain | Serrigar, OS[3] | Pyr | GBSR0273 | -2.9747 | 58.7808 | 98 | | |
| Britain | Settiscarth, OM[6] | Pyr | OMSe0001 | -3.1034 | 59.0505 | 82 | | |
| Britain | Settiscarth, OM[6] | Pyr | OMSe0002 | -3.1034 | 59.0505 | 82 | 8 | |
| Britain | South Gloucestershire[3] | NCE | GBSG0044 | -2.3317 | 51.5617 | 94 | | |
| Britain | Staffordshire[3] | NCE | GBSt0005 | -2.2903 | 52.7233 | 100 | | |
| Britain | Stilligarry, S. Uist[3] | Pyr | GBSU0227 | -7.3708 | 57.3236 | 19 | | |
| Britain | Stilligarry, S. Uist[3] | Pyr | GBSU0310 | -7.3708 | 57.3233 | 106 | | |
| Britain | Tankerness, OM[6] | Pyr | OMTa0001 | -2.8500 | 58.9500 | 83 | | |
| Britain | Tankerness, OM[6] | Pyr | OMTa0002 | -2.8500 | 58.9500 | 83 | | 1 |
| Britain | Tankerness, OM[6] | Pyr | OMTa0003 | -2.8500 | 58.9500 | 83 | | |
| Britain | Tankerness, OM[6] | Pyr | OMTa0004 | -2.8500 | 58.9500 | 83 | | |
| Britain | Tankerness, OM[6] | Pyr | OMTa0005 | -2.8500 | 58.9500 | 83 | | |
| Britain | Tankerness, OM[6] | Pyr | OMTa0006 | -2.8500 | 58.9500 | 83 | | |
| Britain | Tankerness, OM[6] | Pyr | OMTa0007 | -2.8500 | 58.9500 | 83 | | |
| Britain | Tankerness, OM[6] | Pyr | OMTa0008 | -2.8500 | 58.9500 | 277 | | 1 |
| Britain | Tankerness, OM[6] | Pyr | OMTa0009 | -2.8500 | 58.9500 | 83 | | |
| Britain | Thorley, Isle of Wight[3] | Pyr | GBIW0003 | -1.4719 | 50.6889 | 63 | | |
| Britain | Thurvaston, Derbyshire[3] | NCE | GBDe0001 | -1.6369 | 52.9442 | 35 | | |
| Britain | Thurvaston, Derbyshire[3] | NCE | GBDe0002 | -1.6369 | 52.9442 | 36 | | |
| Britain | Tomatin, Inverness[3] | NCE | GBIn0407 | -3.9606 | 57.3364 | 57 | | |
| Britain | Warwickshire[3] | NCE | GBWa0050 | -1.2431 | 52.1272 | 107 | | |
| Britain | Warwickshire[3] | NCE | GBWa0051 | -1.2431 | 52.1272 | 107 | | |
| Britain | Wedholme Flow, Cumbria[3] | NCE | GBCu00B1 | -3.2347 | 54.8678 | 31 | | |
| Britain | Wedholme Flow, Cumbria[3] | NCE | GBCu00B2 | -3.2347 | 54.8678 | 32 | | |
| Britain | Wedholme Flow, Cumbria[3] | NCE | GBCu00D1 | -3.2347 | 54.8678 | 32 | | |
| Britain | West Sussex[3] | NCE | GBWS0011 | 0.3075 | 50.8447 | 117 | | |
| Britain | Wester Ross[3] | Pyr | GBWR0176 | -5.3317 | 57.4819 | 115 | | |

| Britain | Wester Ross[3] | Pyr | GBWR0178 | -5.3317 | 57.4819 | 116 | | |
| Britain | Whiting Bay, Arran[3] | Pyr | GBAr0003 | -5.0992 | 55.4858 | 15 | | |
| Britain | Windwick,OS[6] | Pyr | OSWW0006 | -2.9407 | 58.7668 | 98 | | 1 |
| Britain | Windwick,OS[6] | Pyr | OSWW0007 | -2.9407 | 58.7668 | 98 | | |
| Britain | Windwick,OS[6] | Pyr | OSWW0008 | -2.9407 | 58.7668 | 98 | | |
| Britain | Windwick,OS[6] | Pyr | OSWW0009 | -2.9407 | 58.7668 | 98 | 8 | |
| Britain | Windwick,OS[6] | Pyr | OSWW0010 | -2.9407 | 58.7668 | 98 | | |
| Britain | Windwick,OS[6] | Pyr | OSWW0011 | -2.9407 | 58.7668 | 98 | | |
| Britain | Windwick,OS[6] | Pyr | OSWW0012 | -2.9407 | 58.7668 | 98 | | |
| Britain | Windwick,OS[6] | Pyr | OSWW0013 | -2.9407 | 58.7668 | 98 | | |
| Britain | Windwick,OS[6] | Pyr | OSWW0014 | -2.9407 | 58.7668 | 98 | | |
| Britain | Windwick,OS[6] | Pyr | OSWW0015 | -2.9407 | 58.7668 | 98 | | |
| Britain | Windwick,OS[6] | Pyr | OSWW0017 | -2.9407 | 58.7668 | 98 | | |
| Britain | Windwick,OS[6] | Pyr | OSWW0018 | -2.9407 | 58.7668 | 99 | | |
| Britain | Windwick,OS[6] | Pyr | OSWW0019 | -2.9407 | 58.7668 | 98 | | |
| Britain | Windwick,OS[6] | Pyr | OSWW0020 | -2.9407 | 58.7668 | 98 | | |
| Britain | Windwick,OS[6] | Pyr | OSWW0021 | -2.9407 | 58.7668 | 98 | | |
| Britain | Windwick,OS[6] | Pyr | OSWW0024 | -2.9407 | 58.7668 | 98 | | |
| Britain | Windwick,OS[6] | Pyr | OSWW0025 | -2.9407 | 58.7668 | 99 | | |
| Britain | Wokingham[3] | Pyr | GBWk0037 | -0.9039 | 51.3664 | 109 | | |
| Britain | Wokingham[3] | Pyr | GBWk0038 | -0.9039 | 51.3664 | 110 | | |
| Britain | Worcestershire[3] | NCE | GBWo0009 | -2.1247 | 52.3100 | 112 | | |
| Britain | Worcestershire[3] | NCE | GBWo0011 | -2.0294 | 52.4125 | 113 | | |
| Britain | Worcestershire[3] | NCE | GBWo0012 | -2.4133 | 52.2667 | 114 | | |
| Czech Republic | Bohemia[5] | NCE | CZBo0154 | 13.5695 | 49.8642 | 178 | | |
| Czech Republic | Bohemia[5] | NCE | CZBo0155 | 13.5695 | 49.8642 | 179 | | 4,12 |
| Czech Republic | Bohemia[5] | NCE | CZBo0156 | 13.5695 | 49.8642 | 180 | | |
| Czech Republic | Bohemia[5] | NCE | CZBo0238 | 13.5695 | 49.8642 | | 3 | |
| Czech Republic | Bohemia[6] | Balk | CZBo0153 | 13.5695 | 49.8642 | 177 | 3 | |
| Czech Republic | Ceske Jiretin[1] | NCE | CZCJ0001 | 13.5667 | 50.6833 | 181 | | |
| Czech Republic | Decin City, Bohemia[5] | NCE | CZDe0009 | 14.1988 | 50.8059 | 182 | | |
| Czech Republic | Decin City, Bohemia[5] | NCE | CZDe0010 | 14.1988 | 50.8059 | 183 | 3 | |
| Czech Republic | Most district, Bohemia[5] | NCE | CZMo0794 | 13.5377 | 50.6003 | 184 | | |
| Czech Republic | Oleska[5] | NCE | CZOl0039 | 14.9096 | 49.9486 | 44 | 3 | |
| Czech Republic | Srnin Sumava Mts[5] | NCE | CZSS0237 | 13.4755 | 49.0656 | 185 | | |
| Czech Republic | Srnin Sumava Mts[5] | NCE | CZSS0238 | 13.4755 | 49.0656 | 186 | | |
| Czech Republic | Srnin Sumava Mts[6] | NCI | CZSS4767 | 13.4755 | 49.0656 | 187 | | |
| Czech Republic | Srnin Sumava Mts[6] | NCI | CZSS4838 | 13.4755 | 49.0656 | 187 | | |
| Czech Republic | Vltava River, Bohemia[5] | NCE | CZVl0001 | 14.4129 | 48.9114 | 188 | | |
| Czech Republic | Vltava River, Bohemia[5] | NCE | CZVl0002 | 14.4129 | 48.9114 | 188 | 3 | |
| Czech Republic | Vltava River, Bohemia[5] | NCE | CZVl0003 | 14.4129 | 48.9114 | 189 | 3 | |
| Czech Republic | Vltava River, Bohemia[5] | NCE | CZVl0004 | 14.4129 | 48.9114 | 190 | | |
| Czech Republic | Vltava River, Bohemia[6] | NCE | CZVl0005 | 14.4190 | 48.9115 | 191 | 3 | |
| Czech Republic | Vltava River, Bohemia[6] | NCI | CZVl0006 | 14.4190 | 48.9115 | 192 | | |
| Czech Republic | Vltava River, Bohemia[6] | NCI | CZVl0007 | 14.4190 | 48.9115 | 192 | | |
| Czech | Vltava River, | NCI | CZVl0008 | 14.4190 | 48.9115 | 192 | 3 | |

| Republic | Bohemia[6] | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Denmark | Amager[1] | NCE | DKAm0001 | 12.5833 | 55.5833 | 203 | | |
| Denmark | Bornholm[1] | NCE | DKBo0001 | 15.0000 | 55.0333 | 204 | | |
| Denmark | Bornholm[1] | NCE | DKBo0002 | 15.0000 | 55.0333 | 205 | | |
| Denmark | Fyn[1] | NCE | DKFy0001 | 10.8000 | 55.3167 | 206 | | |
| Denmark | Langeland[1] | NCE | DKLa0001 | 10.7167 | 54.9500 | 207 | | |
| Denmark | Mandø[5] | NCE | DKMa0005 | 8.5521 | 55.2772 | 208 | 7 | |
| Denmark | Mandø[5] | NCE | DKMa0006 | 8.5521 | 55.2772 | 209 | 3 | |
| Denmark | Mandø[5] | NCE | DKMa0007 | 8.5521 | 55.2772 | 210 | | |
| Denmark | Viking centre, Ribe[5] | NCE | DKRi0003 | 8.7628 | 55.3270 | 211 | 3 | |
| Finland | Lammi[1] | NCE | FILa0001 | 25.1167 | 61.0667 | 222 | | |
| Finland | Saortu[1] | NCE | FISa0001 | 27.2500 | 61.7333 | 223 | | |
| France | Belle Île, Brittany[3] | Pyr | FRBI0001 | -3.1958 | 47.3375 | 1 | | |
| France | Belle Île, Brittany[3] | Pyr | FRBI0002 | -3.1958 | 47.3375 | 2 | | |
| France | Belle Île, Brittany[3] | Pyr | FRBI0005 | -3.2000 | 47.3708 | 3 | | |
| France | Belle Île, Brittany[3] | Pyr | FRBI0011 | -3.1222 | 47.3083 | 4 | | |
| France | Belle Île, Brittany[3] | Pyr | FRBI00LP | -3.1500 | 47.3500 | 5 | | |
| France | Broualan, Brittany[3] | NCE | FRBr2McD | -1.6250 | 48.4667 | 6 | | |
| France | Cistriere Auvergne[6] | Pyr | FRCi0003 | 3.5408 | 45.4459 | 224 | 8 | |
| France | Divonne Les Bains, Ain[4] | NCI | FRDi3003 | 6.1432 | 46.3568 | 187 | | |
| France | Etang des Balceres[6] | Pyr | FREB0001 | 2.0802 | 42.5525 | 225 | | 1,3 |
| France | Fressenville, Normandie[5] | NCE | FRFr0139 | 0.0012 | 48.9314 | 226 | | |
| France | Lac des Bouillouses[6] | Pyr | FRLB0066 | 1.9923 | 42.5614 | 228 | 8 | 1 |
| France | Lans en Vercors[4] | NCI | FRLV0002 | 5.5891 | 45.1280 | 227 | | |
| France | Limousin[3] | Pyr | FRLi3McD | 2.1333 | 45.7000 | 7 | | |
| France | Morlaix[1] | NCE | FRMo0001 | -3.8333 | 48.5833 | 229 | | |
| France | Nexon[6] | Pyr | FRNe0004 | 1.2500 | 45.7500 | 5 | 10 | |
| France | Paimpont Broceliande[6] | Pyr | FRPa0001 | -2.2795 | 48.0013 | 231 | | |
| France | Pas de Calais[6] | Pyr | FRPC0001 | 2.3021 | 50.5660 | 231 | | |
| France | Pont Plancoet, Bretagne[5] | NCE | FRPP5540 | -4.1987 | 48.6510 | 229 | 8 | 4,6 |
| France | Pont Saint Marco, Nord[5] | NCE | FRSM0001 | -2.0403 | 48.6144 | 232 | | |
| Germany | Anholt, Saxony[5] | NCE | DESa0003 | 11.3955 | 52.4779 | 201 | 6 | |
| Germany | Anholt, Saxony[5] | NCE | DESa0005 | 11.3955 | 52.4779 | 202 | | |
| Germany | Beltingharder Koog[5] | NCE | DEBK0001 | 8.7846 | 54.6756 | 193 | | |
| Germany | Eberswalde[5] | NCE | DEEb3996 | 13.8109 | 52.8331 | 195 | 3 | 4,13 |
| Germany | Entin[5] | NCE | DEEn0005 | 10.6038 | 54.1359 | 196 | | |
| Germany | Hartz Mts[1] | NCE | DEHM0001 | 10.6667 | 51.7500 | 197 | | |
| Germany | Norderoor[5] | NCE | DENN0002 | 9.1759 | 54.5896 | 198 | 3 | |
| Germany | Norderoor[5] | NCE | DENN0003 | 9.1759 | 54.5896 | 199 | 3 | |
| Germany | Oetisheim[6] | NCI | DEOe0001 | 8.7915 | 48.9708 | 200 | | |
| Ireland | Adare, Limerick[3] | Pyr | IELi0058 | -8.8342 | 52.5842 | 141 | | |
| Ireland | Adare, Limerick[3] | Pyr | IELi0059 | -8.8342 | 52.5842 | 142 | | |
| Ireland | Adare, Limerick[3] | Pyr | IELi0060 | -8.8342 | 52.5842 | 143 | | |
| Ireland | Adare, Limerick[3] | Pyr | IELi0068 | -8.8342 | 52.5842 | 127 | | |
| Ireland | Adare, Limerick[3] | Pyr | IELi0069 | -8.7692 | 52.6175 | 127 | | |
| Ireland | Adare, Limerick[3] | Pyr | IELi0080 | -8.7692 | 52.6175 | 127 | | |
| Ireland | Adare, Limerick[3] | Pyr | IELi0082 | -8.7692 | 52.6175 | 144 | | |
| Ireland | Aran Island[6] | Pyr | IEAI0001 | -8.5286 | 54.9934 | 127 | 8 | |
| Ireland | Ballinamore, Leitrim[3] | Pyr | IELe0001 | -7.7861 | 53.9583 | 140 | | |
| Ireland | Ballinamore, Leitrim[3] | Pyr | IELe0007 | -7.7861 | 53.9583 | 127 | | |
| Ireland | Ballinamore, Leitrim[3] | Pyr | IELe0008 | -7.7689 | 53.9911 | 127 | | |
| Ireland | Ballinamore, | Pyr | IELe0018 | -7.7733 | 53.9867 | 127 | | |

| | Leitrim[3] | | | | | | |
|---|---|---|---|---|---|---|---|
| Ireland | Ballinamore, Leitrim[3] | Pyr | IELe0019 | -7.7733 | 53.9867 | 127 | |
| Ireland | Ballycastle, Mayo[3] | Pyr | IEMa0032 | -9.3167 | 54.2333 | 147 | |
| Ireland | Ballycastle, Mayo[3] | Pyr | IEMa0037 | -9.3167 | 54.2333 | 148 | |
| Ireland | Camolin, Wexford[1] | Pyr | IECa0001 | -6.4167 | 52.5833 | 244 | |
| Ireland | Camolin, Wexford[3] | Pyr | IEWe0036 | -6.4167 | 52.5833 | 152 | |
| Ireland | Camolin, Wexford[3] | Pyr | IEWe0054 | -6.4167 | 52.5833 | 127 | |
| Ireland | Camolin, Wexford[3] | Pyr | IEWe0055 | -6.4167 | 52.5833 | 127 | |
| Ireland | Carlingford, Louth[3] | Pyr | IELo011a | -6.1833 | 54.0333 | 145 | |
| Ireland | Carlingford, Louth[3] | Pyr | IELo011b | -6.1833 | 54.0333 | 145 | |
| Ireland | Carlingford, Louth[3] | Pyr | IELo011c | -6.1833 | 54.0333 | 145 | |
| Ireland | Castlebridge[1] | Pyr | IECt0001 | -6.5000 | 52.4167 | 246 | |
| Ireland | Cloghan[1] | Pyr | IECl0001 | -7.7500 | 53.2500 | 245 | |
| Ireland | Durrow, Laois[3] | Pyr | IELa0LSa | -7.5000 | 52.7833 | 138 | |
| Ireland | Durrow, Laois[3] | Pyr | IELa0LSb | -7.5000 | 52.7833 | 139 | |
| Ireland | Durrow, Laois[3] | Pyr | IELa0LSc | -7.5000 | 52.7833 | 127 | |
| Ireland | Glendalough, Wicklow[3] | Pyr | IEWi00Qa | -6.3833 | 53.0000 | 127 | |
| Ireland | Glendalough, Wicklow[3] | Pyr | IEWi00Qb | -6.3833 | 53.0000 | 154 | |
| Ireland | Glendalough, Wicklow[3] | Pyr | IEWi00Qc | -6.3833 | 53.0000 | 127 | |
| Ireland | Greysteel, Derry[3] | Pyr | IEDy0001 | -7.2500 | 55.0000 | 127 | |
| Ireland | Greysteel, Derry[3] | Pyr | IEDy0003 | -7.2500 | 55.0000 | 128 | |
| Ireland | Greysteel, Derry[3] | Pyr | IEDy0004 | -7.2500 | 55.0000 | 127 | |
| Ireland | Greysteel, Derry[3] | Pyr | IEDy0005 | -7.2500 | 55.0000 | 128 | |
| Ireland | Greysteel, Derry[3] | Pyr | IEDy0006 | -7.2500 | 55.0000 | 128 | |
| Ireland | Greysteel, Derry[3] | Pyr | IEDy0007 | -7.2500 | 55.0000 | 128 | |
| Ireland | Greysteel, Derry[3] | Pyr | IEDy0008 | -7.2500 | 55.0000 | 128 | |
| Ireland | Greysteel, Derry[3] | Pyr | IEDy0009 | -7.2500 | 55.0000 | 128 | |
| Ireland | Greysteel, Derry[3] | Pyr | IEDy0010 | -7.2500 | 55.0000 | 128 | |
| Ireland | Greysteel, Derry[3] | Pyr | IEDy0011 | -7.2500 | 55.0000 | 128 | |
| Ireland | Greysteel, Derry[3] | Pyr | IEDy0012 | -7.2500 | 55.0000 | 128 | |
| Ireland | Gweedore, Donegal[3] | Pyr | IEGW0001 | -8.2300 | 55.0500 | 133 | |
| Ireland | Gweedore, Donegal[3] | Pyr | IEGW0026 | -8.2300 | 55.0500 | 133 | |
| Ireland | Gweedore, Donegal[3] | Pyr | IEGW005a | -8.2300 | 55.0500 | 133 | |
| Ireland | Gweedore, Donegal[3] | Pyr | IEGW005b | -8.2300 | 55.0500 | 133 | |
| Ireland | Kill, Waterford[3] | Pyr | IEWa0001 | -7.3333 | 52.1750 | 150 | |
| Ireland | Kill, Waterford[3] | Pyr | IEWa0002 | -7.3333 | 52.1750 | 127 | |
| Ireland | Kill, Waterford[3] | Pyr | IEWa0003 | -7.3333 | 52.1750 | 136 | |
| Ireland | Kill, Waterford[3] | Pyr | IEWa0004 | -7.3333 | 52.1750 | 150 | |
| Ireland | Kill, Waterford[3] | Pyr | IEWa0005 | -7.3333 | 52.1750 | 136 | |
| Ireland | Killala, Mayo[3] | Pyr | IEMa0020 | -9.2500 | 54.1667 | 146 | |
| Ireland | Killala, Mayo[3] | Pyr | IEMa0025 | -9.2500 | 54.1667 | 146 | |
| Ireland | Killarney, Kerry[3] | Pyr | IEKe0001 | -9.5333 | 52.0556 | 127 | |
| Ireland | Killarney, Kerry[3] | Pyr | IEKe0002 | -9.5333 | 52.0556 | 134 | |
| Ireland | Killarney, Kerry[3] | Pyr | IEKe0004 | -9.5333 | 52.0556 | 127 | |
| Ireland | Killarney, Kerry[3] | Pyr | IEKe0005 | -9.5333 | 52.0556 | 127 | |
| Ireland | Killarney, Kerry[3] | Pyr | IEKeP3kY | -9.5333 | 52.0556 | 135 | |
| Ireland | Kinvarra, Galway[3] | Pyr | IEGa0039 | -8.8667 | 53.1500 | 129 | |
| Ireland | Kinvarra, Galway[3] | Pyr | IEGa0046 | -8.8667 | 53.1500 | 130 | |
| Ireland | Kinvarra, Galway[3] | Pyr | IEGa0047 | -8.9667 | 53.1333 | 127 | |
| Ireland | Kinvarra, Galway[3] | Pyr | IEGa0050 | -8.9667 | 53.1333 | 131 | |
| Ireland | Kinvarra, Galway[3] | Pyr | IEGa0057 | -8.9667 | 53.1333 | 132 | |
| Ireland | Mountcharles, Donegal[3] | Pyr | IEDN0001 | -8.3000 | 54.6500 | 125 | |
| Ireland | Mountcharles, Donegal[3] | Pyr | IEDN0009 | -8.3000 | 54.6500 | 126 | |
| Ireland | Offaly, Cloghan[3] | Pyr | IEOf0019 | -7.7500 | 53.2500 | 149 | |

| Ireland | Offaly, Cloghan[3] | Pyr | IEOf0026 | -7.7500 | 53.2500 | 127 | | |
|---|---|---|---|---|---|---|---|---|
| Ireland | Offaly, Cloghan[3] | Pyr | IEOf0048 | -7.7500 | 53.2500 | 149 | | |
| Ireland | Rathangan, Kildare[3] | Pyr | IEKi0084 | -6.9778 | 53.2797 | 127 | | |
| Ireland | Rathangan, Kildare[3] | Pyr | IEKi0085 | -6.9778 | 53.2797 | 127 | | |
| Ireland | Rathangan, Kildare[3] | Pyr | IEKi0086 | -6.9778 | 53.2797 | 136 | | |
| Ireland | Rathangan, Kildare[3] | Pyr | IEKi0087 | -6.9778 | 53.2797 | 127 | | |
| Ireland | Rathangan, Kildare[3] | Pyr | IEKi0089 | -6.9778 | 53.2797 | 137 | | |
| Ireland | Slane[1] | Pyr | IESl0001 | -6.5000 | 53.7500 | 127 | | |
| Ireland | Tintern Abbey, Wexford[3] | Pyr | IEWe0005 | -6.8417 | 52.2333 | 151 | | |
| Ireland | Tintern Abbey, Wexford[3] | Pyr | IEWe0008 | -6.8417 | 52.2333 | 127 | | |
| Ireland | Tintern Abbey, Wexford[3] | Pyr | IEWe0018 | -6.8417 | 52.2333 | 151 | | |
| Ireland | Tintern Abbey, Wexford[3] | Pyr | IEWe0031 | -6.8417 | 52.2333 | 127 | | |
| Ireland | Tintern Abbey, Wexford[3] | Pyr | IEWe0033 | -6.8417 | 52.2333 | 151 | | |
| Ireland | Tintern Abbey, Wexford[3] | Pyr | IEWe020a | -6.8417 | 52.2333 | 127 | | |
| Ireland | Tintern Abbey, Wexford[3] | Pyr | IEWe023a | -6.8417 | 52.2333 | 153 | | |
| Ireland | Tintern Abbey, Wexford[3] | Pyr | IEWe023b | -6.8417 | 52.2333 | 127 | | |
| Ireland | Whiting Bay, Cork[1] | Pyr | IEWB0001 | -7.8333 | 51.8333 | 247 | | |
| Ireland | Whiting Bay, Cork[1] | Pyr | IEWB0002 | -7.8333 | 51.8333 | 248 | | |
| Ireland | Whiting Bay, Cork[1] | Pyr | IEWB0003 | -7.8333 | 51.8333 | 118 | | |
| Ireland | Whiting Bay, Cork[3] | Pyr | IECo0095 | -7.8283 | 51.9494 | 118 | | |
| Ireland | Whiting Bay, Cork[3] | Pyr | IECo0098 | -7.8283 | 51.9494 | 118 | | |
| Ireland | Whiting Bay, Cork[3] | Pyr | IECo0099 | -7.8283 | 51.9494 | 118 | | |
| Ireland | Whiting Bay, Cork[3] | Pyr | IECo0100 | -7.8283 | 51.9494 | 119 | | |
| Ireland | Whiting Bay, Cork[3] | Pyr | IECo0101 | -7.8283 | 51.9494 | 118 | | |
| Ireland | Whiting Bay, Cork[3] | Pyr | IECo0102 | -7.8283 | 51.9494 | 120 | | |
| Ireland | Whiting Bay, Cork[3] | Pyr | IECo0103 | -7.8283 | 51.9494 | 121 | | |
| Ireland | Whiting Bay, Cork[3] | Pyr | IECo0104 | -7.8283 | 51.9494 | 122 | | |
| Ireland | Whiting Bay, Cork[3] | Pyr | IECo0105 | -7.8283 | 51.9494 | 123 | | |
| Ireland | Whiting Bay, Cork[3] | Pyr | IECo0108 | -7.8283 | 51.9494 | 124 | | |
| Italy | Abruzzo[1] | NCI | ITAb0001 | 14.0000 | 42.0000 | 249 | | |
| Italy | Abruzzo[1] | NCI | ITAb0002 | 14.0000 | 42.0000 | 250 | | |
| Italy | Camigliatello, Calabria[4] | SIta | ITCa5342 | 16.4460 | 39.3386 | 251 | 11 | 10 |
| Italy | Castelfranco, Toscana[4] | NCI | ITTo1578 | 11.5501 | 43.6176 | 262 | | |
| Italy | La Sila, Calabria[4] | SIta | ITSi0011 | 16.4911 | 39.3522 | 258 | | 11 |
| Italy | La Sila, Calabria[4] | SIta | ITSi0017 | 16.4911 | 39.3522 | 259 | 11 | 4,9 |
| Italy | La Sila, Calabria[4] | SIta | ITSi0021 | 16.4911 | 39.3522 | 260 | 11 | 4 |
| Italy | Laghi di Ceretto[2] | Iber | ITLC0001 | 10.2426 | 44.2978 | 218 | 5 | |
| Italy | Majella Mts, Abruzzo[4] | NCI | ITMa00NT | 13.9284 | 42.2854 | 250 | | |
| Italy | Majella Mts, Abruzzo[4] | NCI | ITMa0175 | 14.1157 | 42.0834 | 254 | | 4,5 |
| Italy | Majella Mts, Abruzzo[4] | NCI | ITMa0176 | 14.1157 | 42.0834 | 250 | | 4 |
| Italy | Majella Mts, Abruzzo[4] | NCI | ITMa0177 | 14.1157 | 42.0834 | 255 | | |
| Italy | Majella Mts, Abruzzo[4] | NCI | ITMa0178 | 14.1157 | 42.0834 | 250 | | |
| Italy | Majella Mts, Abruzzo[4] | NCI | ITMa0179 | 14.1157 | 42.0834 | 254 | 5 | |
| Italy | Prasota, Mazzo[4] | NCI | ITPr0001 | 10.2480 | 46.2870 | 256 | | |
| Italy | Prasota, Mazzo[4] | NCE | ITPr0004 | 10.2480 | 46.2870 | 257 | 2 | |
| Italy | San Carlo di Cese[4] | NCI | ITGe0001 | 8.8320 | 44.4773 | 252 | | |

| Italy | Trento[1] | NCI | ITTr0001 | 11.8333 | 46.2500 | 263 | | |
|---|---|---|---|---|---|---|---|---|
| Italy | Trento[1] | NCE | ITTr0002 | 11.8333 | 46.2500 | 264 | | |
| Italy | Valbrevenna, Genova[4] | NCI | ITGe0003 | 9.0648 | 44.5553 | 253 | | |
| Italy | Valle Aosta, Torgnon[4] | NCE | ITTg0049 | 7.5712 | 45.8072 | 261 | | |
| Lithuania | Vilnius[1] | NCE | LTVi0001 | 25.3167 | 54.6667 | 265 | | |
| Lithuania | Vilnius[1] | NCE | LTVi0002 | 25.3167 | 54.6667 | 266 | | |
| Macedonia | Jakupica Mts[6] | Balk | MKJa1247 | 21.4189 | 41.6891 | 269 | | |
| Macedonia | Jakupica Mts[6] | Balk | MKJa9222 | 21.4189 | 41.6891 | 270 | | |
| Macedonia | Mount Galicica[6] | Balk | MKMG0001 | 20.8508 | 41.1015 | 271 | | 4 |
| Macedonia | Pelister Mts[1] | Balk | MKPe0001 | 21.1667 | 41.0000 | 272 | | |
| Montenegro | Lovćen[6] | Balk | MELo6362 | 18.8188 | 42.4058 | 267 | | |
| Montenegro | Mount Komovi[6] | Balk | MEMK0001 | 19.6581 | 42.6947 | 268 | 12 | 4 |
| Netherlands | Ameland, Friesland[5] | NCE | NLFr0017 | 5.7960 | 53.4453 | 10 | | |
| Netherlands | Ameland, Friesland[5] | NCE | NLFr0040 | 5.7960 | 53.4453 | 10 | | |
| Netherlands | Boxtel[1] | NCE | NLBo0001 | 5.3333 | 51.6000 | 273 | | |
| Netherlands | Callantsoog[5] | NCE | NLCa0002 | 4.6937 | 52.8303 | 274 | | |
| Netherlands | Dieren, Gelderland[5] | NCE | NLDi0001 | 6.0815 | 52.0479 | 10 | | |
| Netherlands | Diessen[5] | NCE | NLDi0062 | 5.1752 | 51.4748 | 201 | | |
| Netherlands | Overijssel[5] | NCE | NLOv0042 | 6.4570 | 52.4953 | 275 | | |
| Netherlands | Wageningen[1] | NCE | NLWa0001 | 5.6667 | 51.9667 | 276 | | |
| Netherlands | Zeeuws Vlaanderen[5] | NCE | NLZV0044 | 5.7221 | 52.7181 | 273 | 8 | |
| Norway | Askland[3] | NCE | NOAs0301 | 8.4500 | 58.6000 | 155 | | |
| Poland | Blizocin[1] | NCE | PLBl0001 | 22.2667 | 51.6000 | 278 | | |
| Russia | Brjansk[1] | NCE | RUBr0001 | 34.0000 | 52.3333 | 280 | | |
| Russia | Brjansk[1] | NCE | RUBr0002 | 34.0000 | 52.3333 | 281 | | |
| Russia | Lake Baikal, Siberia[1] | NCE | RULB0001 | 108.0000 | 53.6667 | 282 | | |
| Russia | Novosibirsk, Siberia[1] | NCE | RUNo0002 | 83.1000 | 54.8167 | 284 | | |
| Russia | Novosibirsk, Siberia[1] | NCE | RUNo0003 | 83.1000 | 54.8167 | 285 | | |
| Russia | Novosibirsk, Siberia[2] | NCE | RUNo0001 | 82.7791 | 55.2014 | 283 | | |
| Russia | Pertozero[1] | NCE | RUPe0001 | 34.0000 | 62.0833 | 286 | | |
| Russia | Tambov[1] | NCE | RUTa0001 | 42.2500 | 51.9167 | 287 | | |
| Russia | Tambov[1] | NCE | RUTa0002 | 42.2500 | 51.9167 | 288 | | |
| Serbia | Jastrebac[6] | Balk | RSJa7390 | 20.5871 | 44.0936 | 279 | | |
| Serbia | Mount Kopaonik[6] | Balk | RSMK1078 | 20.8936 | 43.1632 | 270 | | |
| Serbia | Mount Kopaonik[6] | Balk | RSMK7008 | 20.8936 | 43.1632 | 270 | 13 | |
| Serbia | Valjevo[6] | Balk | RSVa7855 | 19.8828 | 44.2463 | 270 | | |
| Slovakia | Bratislava[5] | NCE | SKBr5733 | 17.1077 | 48.1479 | 161 | 3 | |
| Slovakia | Bratislava[5] | NCE | SKBr5736 | 17.1077 | 48.1479 | 302 | | |
| Slovakia | Bratislava[5] | NCE | SKBr5743 | 17.1077 | 48.1479 | 161 | 3 | |
| Slovakia | Bratislava[6] | Balk | SKBr5730 | 17.1077 | 48.1479 | 301 | | |
| Slovakia | Bratislava[6] | Balk | SKBr5734 | 17.1077 | 48.1479 | 301 | 3 | |
| Slovakia | Bratislava[6] | Balk | SKBr5735 | 17.1077 | 48.1479 | 301 | | |
| Slovakia | Kezmarok, Jezersko[5] | NCE | SKKe0418 | 20.3711 | 49.2887 | 44 | 3 | |
| Slovakia | Kezmarok, Jezersko[5] | NCE | SKKe0419 | 20.3711 | 49.2887 | 97 | 3 | 4 |
| Slovenia | Nova Gorica, Anhovo[4] | NCI | SING0001 | 13.6571 | 45.9447 | 299 | | 4 |
| Slovenia | Postjma[4] | NCI | SIPo0001 | 14.2082 | 45.7811 | 300 | 14 | 4 |
| Spain | Cantabria[6] | Pyr | ESEM0069 | -3.4503 | 43.1423 | 213 | 8 | 6 |
| Spain | Navarra[6] | Pyr | ESNa0047 | -1.6455 | 43.1757 | 214 | | |
| Spain | Navarra[6] | Pyr | ESNa0861 | -1.6455 | 43.1757 | 127 | | |
| Spain | Navarra[6] | Pyr | ESNa1131 | -1.6455 | 43.1757 | 127 | | |
| Spain | Navarra[6] | Pyr | ESNa1286 | -1.6455 | 43.1757 | 214 | | |
| Spain | Picos de Europa[6] | Iber | ESPE0047 | -4.9997 | 43.1049 | 215 | 9 | 4,6 |

| Spain | Picos de Europa[6] | Iber | ESPE0057 | -4.9997 | 43.1049 | 216 | 8 | 2 |
|---|---|---|---|---|---|---|---|---|
| Spain | Pyrenees[6] | Pyr | ESBa9709 | 1.1136 | 42.5650 | 212 | | |
| Spain | Rascafria[1] | Iber | ESRa0001 | -3.8833 | 40.9000 | 217 | | |
| Spain | Rascafria[1] | Iber | ESRa0002 | -3.8833 | 40.9000 | 218 | | |
| Spain | Rascafria[1] | Iber | ESRa0003 | -3.8833 | 40.9000 | 219 | | |
| Spain | Rascafria[6] | Iber | ESRa0640 | -3.8794 | 40.9036 | 220 | 8 | 4,6,7,8 |
| Spain | Rascafria[6] | Balk | ESRa3448 | -3.8794 | 40.9036 | 221 | | |
| Spain | Rascafria[6] | Iber | ESRa5939 | -3.8794 | 40.9036 | 218 | 8 | 4 |
| Sweden | Jamj[1] | NCE | SEJa0001 | 15.8667 | 56.1667 | 292 | | |
| Sweden | Jamj[1] | NCE | SEJa0002 | 15.8667 | 56.1667 | 293 | | |
| Sweden | Oland[1] | NCE | SEOl0001 | 17.0833 | 57.2667 | 294 | | |
| Sweden | Oland[1] | NCE | SEOl0002 | 17.0833 | 57.2667 | 295 | | |
| Sweden | Revinge[1] | NCE | SERe0001 | 14.3333 | 55.5833 | 296 | | |
| Sweden | Revinge[1] | NCE | SERe0002 | 14.3333 | 55.5833 | 297 | | |
| Sweden | Tingstade, Gotland[1] | NCE | SEGo0003 | 18.6000 | 57.7000 | 291 | | |
| Sweden | Uppsala[1] | NCE | SEUp0001 | 17.6667 | 59.7500 | 298 | | |
| Sweden | Vastergarn, Gotland[1] | NCE | SEGo0001 | 18.1667 | 57.4667 | 289 | | |
| Sweden | Vastergarn, Gotland[1] | NCE | SEGo0002 | 18.1667 | 57.4667 | 290 | | |
| Switzerland | Bassin, Vaud[1] | NCE | CHBa0001 | 6.6500 | 46.5333 | 166 | | |
| Switzerland | Bassins, Vaud[4] | NCE | CHBa5756 | 6.2311 | 46.4628 | 167 | 3 | |
| Switzerland | Bassins, Vaud[6] | NCE | CHBa5698 | 6.2311 | 46.4628 | 315 | | |
| Switzerland | Bassins, Vaud[6] | Balk | CHBa5712 | 6.2311 | 46.4628 | 316 | | |
| Switzerland | Bretolet, Valais[4] | NCI | CHBr5421 | 6.8652 | 46.1689 | 168 | 5 | 4,6 |
| Switzerland | Chablais Cudrefin[4] | NCE | CHCu7581 | 7.0266 | 46.9593 | 172 | | |
| Switzerland | Chalet a Gobet[4] | NCE | CHCG5272 | 6.6927 | 46.5646 | 170 | | |
| Switzerland | Chalet des Enfants[4] | Balk | CHCE0889 | 6.6644 | 46.5742 | 169 | 3 | 4 |
| Switzerland | Champmartin[4] | NCE | CHCh7622 | 6.9974 | 46.9327 | 171 | 3 | |
| Switzerland | Gletterens[4] | NCE | CHGl7628 | 6.9361 | 46.8927 | 173 | 3 | |
| Switzerland | Ostende Chevroux[4] | NCE | CHOC7576 | 6.9178 | 46.8943 | 173 | 3 | |
| Switzerland | Ostende Chevroux[4] | NCE | CHOC7583 | 6.9178 | 46.8943 | 174 | 3 | 4,12 |
| Switzerland | Pont de Nant, Vaud[6] | NCI | CHPN5442 | 7.0943 | 46.2491 | 317 | | |
| Switzerland | Val d'Illiez[4] | NCE | CHVI4747 | 6.8927 | 46.2043 | 175 | | |
| Switzerland | Val d'Illiez[4] | NCI | CHVI4748 | 6.8927 | 46.2043 | 176 | 3 | |
| Turkey | Anatolia[1] | Out | TRAr0001 | 31.7238 | 40.8988 | 322 | | |
| Turkey | Anatolia[6] | Out | TRAr6106 | 31.7238 | 40.8988 | 321 | | |
| Turkey | Anatolia[6] | Out | TRSA6079 | 31.7238 | 40.8988 | 320 | | |
| Turkey | Strandzha Mts[1] | Balk | TRSM0001 | 27.6833 | 41.7500 | 303 | | |
| Turkey | Strandzha Mts[1] | Balk | TRSM0002 | 27.6833 | 41.7500 | 304 | | |
| Ukraine | Cherkassy[1] | NCE | UACh0001 | 31.5000 | 49.7167 | 305 | | |
| Ukraine | Jaduty[5] | NCE | UAJa0043 | 32.3167 | 51.3667 | 306 | 3 | |
| Ukraine | Kanev[5] | NCE | UAKa0024 | 31.8349 | 49.6928 | 307 | | 4 |
| Ukraine | Kanev[5] | NCE | UAKa0025 | 31.8349 | 49.6928 | 308 | 3 | |
| Ukraine | Kanev[5] | NCE | UAKa0250 | 31.8349 | 49.6928 | 308 | | |
| Ukraine | L'Vov[5] | NCE | UALV0255 | 31.0833 | 47.9000 | 309 | 3 | |
| Ukraine | Tishki[5] | NCE | UATi0266 | 33.1109 | 50.1076 | 310 | | |
| Ukraine | Vinnitsa[5] | NCE | DEBK0020 | 8.7846 | 54.6756 | 194 | | |
| Ukraine | Vinnitsa[5] | NCE | UAVi0253 | 28.5623 | 49.2307 | 311 | 3 | |
| Ukraine | Vinnitsa[5] | NCE | UAVi0254 | 28.5623 | 49.2307 | 312 | 3 | |
| Ukraine | Volyn[5] | NCE | DEBK0021 | 8.7846 | 54.6756 | 193 | | |
| Ukraine | Volyn[5] | NCE | UAVo0256 | 24.8567 | 51.1240 | 313 | | |
| Ukraine | Zhitomer[5] | NCE | UAZh0257 | 28.6149 | 50.2617 | 314 | | |

[1] Mascheretti et al. (2003)

[2] Ohdachi et al. (2006)

[3] Searle et al. (2009)

[4] Vega et al. (2010a)

| | |
|---|---|
| [5]Vega et al. (2010b) | |
| [6]This study | |
| Pyr = Pyrenean | |
| NCE = North-Central European | |
| NCI = North-Central Italian | |
| Balk = Balkan | |
| SIta = South Italian | |
| Iber = Iberia | |
| Out = Outgroup | |
| OM = Orkney Mainland | |
| OS = Orkney South Ronaldsay | |
| OW = Orkney Westray | |

**Appendix 4.** List of mandible and skull samples for geometric morphometric analysis.

Mandible and skull datasets for geometric morphometric analysis.

| Country | Code | Phylo group | Longitude (Decimal) | Latitude (Decimal) | Mand-ible | Skull | Group | Island Size Category |
|---|---|---|---|---|---|---|---|---|
| Austria | ATDo1611 | Balk | 16.6413 | 47.8957 | x | x | | |
| Austria | ATDo1612 | NCE | 16.6413 | 47.8957 | x | x | | |
| Austria | ATUmA139 | NCE | 10.9277 | 47.1361 | x | | | |
| Bosnia Herzegovina | BAKu2517 | Balk | 17.3207 | 44.0022 | x | x | | |
| Bosnia Herzegovina | BAOs5670 | Balk | 16.2887 | 44.2397 | x | x | | |
| Bosnia Herzegovina | BAZe4239 | Balk | 18.3889 | 43.3949 | x | x | | |
| Britain | GBCa1 | NCE | -0.8926 | 53.9647 | x | x | | |
| Britain | GBCH1 | NCE | -0.9104 | 54.1209 | x | x | | |
| Britain | GBDrG140 | NCE | -4.4893 | 57.3090 | x | | | |
| Britain | GBHe1 | NCE | -1.0577 | 53.9432 | x | x | | |
| Britain | OMHa10 | OM | -3.1902 | 59.0337 | x | x | Main1 | Med |
| Britain | OMHa11 | OM | -3.1902 | 59.0337 | x | x | Main1 | Med |
| Britain | OMHa12 | OM | -3.1902 | 59.0337 | x | x | Main1 | Med |
| Britain | OMHa3BM | OM | -3.1902 | 59.0337 | x | x | Main1 | Med |
| Britain | OMHa3CM | OM | -3.1902 | 59.0337 | x | x | Main1 | Med |
| Britain | OMHa3M | OM | -3.1902 | 59.0337 | x | x | Main1 | Med |
| Britain | OMHa4 | OM | -3.1902 | 59.0337 | x | x | Main1 | Med |
| Britain | OMHa6 | OM | -3.1902 | 59.0337 | x | x | Main1 | Med |
| Britain | OMHa9F | OM | -3.1902 | 59.0337 | x | x | Main1 | Med |
| Britain | OMHa9Fb | OM | -3.1902 | 59.0337 | x | | Main1 | Med |
| Britain | OMHa9M | OM | -3.1902 | 59.0337 | x | x | Main1 | Med |
| Britain | OMHo15 | OM | -3.0676 | 58.9464 | x | x | Main2 | Med |
| Britain | OMSa1 | OM | -3.2972 | 59.0483 | x | x | Main1 | Med |
| Britain | OMSa2 | OM | -3.2972 | 59.0483 | x | x | Main1 | Med |
| Britain | OMSO1 | OM | -2.9500 | 58.9500 | x | | Main2 | Med |
| Britain | OMSO10F | OM | -2.9500 | 58.9500 | x | x | Main2 | Med |
| Britain | OMSO11 | OM | -2.9500 | 58.9500 | x | x | Main2 | Med |
| Britain | OMSO12 | OM | -2.9500 | 58.9500 | x | x | Main2 | Med |
| Britain | OMSO13 | OM | -2.9500 | 58.9500 | x | x | Main2 | Med |
| Britain | OMSO15 | OM | -2.9500 | 58.9500 | x | x | Main2 | Med |
| Britain | OMSO16 | OM | -2.9500 | 58.9500 | x | x | Main2 | Med |
| Britain | OMSO17 | OM | -2.9500 | 58.9500 | x | x | Main2 | Med |
| Britain | OMSO18 | OM | -2.9500 | 58.9500 | x | x | Main2 | Med |
| Britain | OMSO2 | OM | -2.9500 | 58.9500 | x | | Main2 | Med |
| Britain | OMSO20 | OM | -2.9500 | 58.9500 | x | x | Main2 | Med |
| Britain | OMSO21 | OM | -2.9500 | 58.9500 | x | x | Main2 | Med |
| Britain | OMSO24 | OM | -2.9500 | 58.9500 | x | x | Main2 | Med |
| Britain | OMSO25 | OM | -2.9500 | 58.9500 | x | x | Main2 | Med |
| Britain | OMSO26 | OM | -2.9500 | 58.9500 | x | x | Main2 | Med |
| Britain | OMSO28 | OM | -2.9500 | 58.9500 | x | x | Main2 | Med |
| Britain | OMSO29 | OM | -2.9500 | 58.9500 | x | x | Main2 | Med |
| Britain | OMSO3 | OM | -2.9500 | 58.9500 | x | x | Main2 | Med |
| Britain | OMSO30 | OM | -2.9500 | 58.9500 | x | x | Main2 | Med |
| Britain | OMSO31 | OM | -2.9500 | 58.9500 | x | x | Main2 | Med |
| Britain | OMSO35 | OM | -2.9500 | 58.9500 | x | x | Main2 | Med |
| Britain | OMSO36 | OM | -2.9500 | 58.9500 | x | x | Main2 | Med |
| Britain | OMSO37 | OM | -2.9500 | 58.9500 | x | x | Main2 | Med |

| Britain | OMSO38 | OM | -2.9500 | 58.9500 | x | x | Main2 | Med |
|---------|--------|----|---------|---------|---|---|-------|-----|
| Britain | OMSO4 | OM | -2.9500 | 58.9500 | x | x | Main2 | Med |
| Britain | OMSO5 | OM | -2.9500 | 58.9500 | x | x | Main2 | Med |
| Britain | OMSO6M | OM | -2.9500 | 58.9500 | x | x | Main2 | Med |
| Britain | OMSO7F | OM | -2.9500 | 58.9500 | x | x | Main2 | Med |
| Britain | OMSO8M | OM | -2.9500 | 58.9500 | x | x | Main2 | Med |
| Britain | OMSO9F | OM | -2.9500 | 58.9500 | x | x | Main2 | Med |
| Britain | OMTa10 | OM | -2.8500 | 58.9500 | x | x | Main2 | Med |
| Britain | OMTa2 | OM | -2.8500 | 58.9500 | x | x | Main2 | Med |
| Britain | OMTa3 | OM | -2.8500 | 58.9500 | x | x | Main2 | Med |
| Britain | OMTa4 | OM | -2.8500 | 58.9500 | x | x | Main2 | Med |
| Britain | OMTa5 | OM | -2.8500 | 58.9500 | x | x | Main2 | Med |
| Britain | OMTa6 | OM | -2.8500 | 58.9500 | x | x | Main2 | Med |
| Britain | OMTa7 | OM | -2.8500 | 58.9500 | x | x | Main2 | Med |
| Britain | OMTa8 | OM | -2.8500 | 58.9500 | x | x | Main2 | Med |
| Britain | OMTa9 | OM | -2.8500 | 58.9500 | | x | Main2 | Med |
| Britain | OSGr10 | OS | -2.9167 | 58.8167 | x | x | Grimn | Small |
| Britain | OSGr12 | OS | -2.9167 | 58.8167 | x | x | Grimn | Small |
| Britain | OSGr14 | OS | -2.9167 | 58.8167 | x | x | Grimn | Small |
| Britain | OSGr15 | OS | -2.9167 | 58.8167 | x | x | Grimn | Small |
| Britain | OSGr18 | OS | -2.9167 | 58.8167 | x | x | Grimn | Small |
| Britain | OSGr20 | OS | -2.9167 | 58.8167 | x | x | Grimn | Small |
| Britain | OSGr23 | OS | -2.9167 | 58.8167 | x | x | Grimn | Small |
| Britain | OSGr25 | OS | -2.9167 | 58.8167 | x | x | Grimn | Small |
| Britain | OSGr27 | OS | -2.9167 | 58.8167 | x | x | Grimn | Small |
| Britain | OSGr28 | OS | -2.9167 | 58.8167 | x | x | Grimn | Small |
| Britain | OSGr31 | OS | -2.9167 | 58.8167 | x | x | Grimn | Small |
| Britain | OSGr35 | OS | -2.9167 | 58.8167 | x | x | Grimn | Small |
| Britain | OSGr39 | OS | -2.9167 | 58.8167 | x | x | Grimn | Small |
| Britain | OSGr42 | OS | -2.9167 | 58.8167 | x | x | Grimn | Small |
| Britain | OSGr45 | OS | -2.9167 | 58.8167 | x | x | Grimn | Small |
| Britain | OSGr49 | OS | -2.9167 | 58.8167 | x | x | Grimn | Small |
| Britain | OSGr50 | OS | -2.9167 | 58.8167 | x | x | Grimn | Small |
| Britain | OSGr7 | OS | -2.9167 | 58.8167 | x | | Grimn | Small |
| Britain | OSGr8 | OS | -2.9167 | 58.8167 | x | x | Grimn | Small |
| Britain | OSGr9 | OS | -2.9167 | 58.8167 | x | x | Grimn | Small |
| Britain | OSWW11 | OS | -2.9407 | 58.7668 | x | x | Wind | Small |
| Britain | OSWW12 | OS | -2.9407 | 58.7668 | x | x | Wind | Small |
| Britain | OSWW13 | OS | -2.9407 | 58.7668 | x | x | Wind | Small |
| Britain | OSWW14 | OS | -2.9407 | 58.7668 | x | x | Wind | Small |
| Britain | OSWW15 | OS | -2.9407 | 58.7668 | x | x | Wind | Small |
| Britain | OSWW16 | OS | -2.9407 | 58.7668 | x | | Wind | Small |
| Britain | OSWW17 | OS | -2.9407 | 58.7668 | x | x | Wind | Small |
| Britain | OSWW18 | OS | -2.9407 | 58.7668 | x | x | Wind | Small |
| Britain | OSWW19 | OS | -2.9407 | 58.7668 | x | x | Wind | Small |
| Britain | OSWW20 | OS | -2.9407 | 58.7668 | x | x | Wind | Small |
| Britain | OSWW21 | OS | -2.9407 | 58.7668 | x | x | Wind | Small |
| Britain | OSWW23 | OS | -2.9407 | 58.7668 | x | x | Wind | Small |
| Britain | OSWW24 | OS | -2.9407 | 58.7668 | x | x | Wind | Small |
| Britain | OSWW25 | OS | -2.9407 | 58.7668 | x | x | Wind | Small |
| Britain | OSWW26 | OS | -2.9407 | 58.7668 | x | x | Wind | Small |
| Britain | OSWW28 | OS | -2.9407 | 58.7668 | x | x | Wind | Small |
| Britain | OSWW6 | OS | -2.9407 | 58.7668 | x | x | Wind | Small |
| Britain | OSWW7 | OS | -2.9407 | 58.7668 | x | x | Wind | Small |
| Britain | OSWW8 | OS | -2.9407 | 58.7668 | x | x | Wind | Small |
| Britain | OSWW9 | OS | -2.9407 | 58.7668 | x | | Wind | Small |
| Britain | OWLS1 | OW | -2.9334 | 59.2833 | x | x | Swart | Small |
| Britain | OWLS112 | OW | -2.9334 | 59.2833 | x | x | Swart | Small |
| Britain | OWLS2 | OW | -2.9334 | 59.2833 | x | x | Swart | Small |
| Britain | OWLS24 | OW | -2.9334 | 59.2833 | x | x | Swart | Small |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Britain | OWLS25 | OW | -2.9334 | 59.2833 | x | x | Swart | Small |
| Britain | OWLS36 | OW | -2.9334 | 59.2833 | x | x | Swart | Small |
| Britain | OWLS37 | OW | -2.9334 | 59.2833 | x | x | Swart | Small |
| Britain | OWLS38 | OW | -2.9334 | 59.2833 | x | x | Swart | Small |
| Britain | OWLS55 | OW | -2.9334 | 59.2833 | x | x | Swart | Small |
| Britain | OWLS59 | OW | -2.9334 | 59.2833 | x | x | Swart | Small |
| Britain | OWLS6 | OW | -2.9334 | 59.2833 | x | x | Swart | Small |
| Britain | OWLS60 | OW | -2.9334 | 59.2833 | x | x | Swart | Small |
| Britain | OWLS61 | OW | -2.9334 | 59.2833 | x | x | Swart | Small |
| Britain | OWLS72 | OW | -2.9334 | 59.2833 | x | x | Swart | Small |
| Britain | OWLS73 | OW | -2.9334 | 59.2833 | x | x | Swart | Small |
| Britain | OWLS82 | OW | -2.9334 | 59.2833 | x | x | Swart | Small |
| Britain | OWLS83 | OW | -2.9334 | 59.2833 | x | x | Swart | Small |
| Britain | OWLS84 | OW | -2.9334 | 59.2833 | x | x | Swart | Small |
| Britain | OWLS85 | OW | -2.9334 | 59.2833 | x | x | Swart | Small |
| Britain | OWLS99 | OW | -2.9334 | 59.2833 | x | x | Swart | Small |
| Britain | OWNe102 | OW | -2.8667 | 59.2334 | x | x | Ness | Small |
| Britain | OWNe103 | OW | -2.8667 | 59.2334 | x | x | Ness | Small |
| Britain | OWNe2 | OW | -2.8667 | 59.2334 | x | x | Ness | Small |
| Britain | OWNe26 | OW | -2.8667 | 59.2334 | x | x | Ness | Small |
| Britain | OWNe28 | OW | -2.8667 | 59.2334 | x | x | Ness | Small |
| Britain | OWNe29 | OW | -2.8667 | 59.2334 | x | x | Ness | Small |
| Britain | OWNe30 | OW | -2.8667 | 59.2334 | x | x | Ness | Small |
| Britain | OWNe31 | OW | -2.8667 | 59.2334 | x | x | Ness | Small |
| Britain | OWNe4 | OW | -2.8667 | 59.2334 | x | x | Ness | Small |
| Britain | OWNe41 | OW | -2.8667 | 59.2334 | x | x | Ness | Small |
| Britain | OWNe42 | OW | -2.8667 | 59.2334 | x | x | Ness | Small |
| Britain | OWNe43 | OW | -2.8667 | 59.2334 | x | x | Ness | Small |
| Britain | OWNe47 | OW | -2.8667 | 59.2334 | x | | Ness | Small |
| Britain | OWNe6 | OW | -2.8667 | 59.2334 | x | x | Ness | Small |
| Britain | OWNe62 | OW | -2.8667 | 59.2334 | x | x | Ness | Small |
| Britain | OWNe68 | OW | -2.8667 | 59.2334 | x | x | Ness | Small |
| Britain | OWNe7 | OW | -2.8667 | 59.2334 | x | x | Ness | Small |
| Britain | OWNe75 | OW | -2.8667 | 59.2334 | x | x | Ness | Small |
| Britain | OWNe89 | OW | -2.8667 | 59.2334 | x | x | Ness | Small |
| Britain | OWNe90 | OW | -2.8667 | 59.2334 | x | x | Ness | Small |
| Czech Republic | CZHa9166 | NCI | 15.5607 | 49.6048 | x | x | | |
| Czech Republic | CZSS4767 | NCI | 13.4755 | 49.0656 | x | x | | |
| Czech Republic | CZSS4838 | NCI | 13.4755 | 49.0656 | x | x | | |
| Finland | FIAE1747 | NCE | 19.6118 | 60.2037 | x | x | | |
| Finland | FIAE1760 | NCE | 19.6118 | 60.2037 | x | x | | |
| Finland | FIKu2071 | NCE | 29.4957 | 64.1254 | x | x | | |
| Finland | FISo1773 | NCE | 27.5296 | 63.6695 | x | x | | |
| Finland | FISo1779 | NCE | 27.5296 | 63.6695 | x | x | | |
| Finland | FISo1783 | NCE | 27.5296 | 63.6695 | x | x | | |
| Finland | FISo1785 | NCE | 27.5296 | 63.6695 | x | x | | |
| France | FRAF174 | NCE | 5.6397 | 48.8561 | x | | | |
| France | FRAF175 | NCE | 5.6397 | 48.8561 | x | | | |
| France | FRAF176 | NCE | 5.6397 | 48.8561 | x | | | |
| France | FRAF179 | NCE | 5.6397 | 48.8561 | x | | | |
| France | FRAF180 | NCE | 5.6397 | 48.8561 | x | | | |
| France | FRAF181 | NCE | 5.6397 | 48.8561 | x | | | |
| France | FRAF182 | NCE | 5.6397 | 48.8561 | x | | | |
| France | FRAF183 | NCE | 5.6397 | 48.8561 | x | | | |
| France | FRAF184 | NCE | 5.6397 | 48.8561 | x | | | |
| France | FRAF185 | NCE | 5.6397 | 48.8561 | x | | | |
| France | FRAF186 | NCE | 5.6397 | 48.8561 | x | | | |
| France | FRAF187 | NCE | 5.6397 | 48.8561 | x | | | |

| France | FRAF188 | NCE | 5.6397 | 48.8561 | x | | |
|--------|---------|-----|--------|---------|---|---|---|
| France | FRAF189 | NCE | 5.6397 | 48.8561 | x | | |
| France | FRAF190 | NCE | 5.6397 | 48.8561 | x | | |
| France | FRAF192 | NCE | 5.6397 | 48.8561 | x | | |
| France | FRAF193 | NCE | 5.6397 | 48.8561 | x | | |
| France | FRAF194 | NCE | 5.6397 | 48.8561 | x | | |
| France | FRAF195 | NCE | 5.6397 | 48.8561 | x | | |
| France | FRAF196A | NCE | 5.6397 | 48.8561 | x | | |
| France | FRAF196B | NCE | 5.6397 | 48.8561 | x | | |
| France | FRAF197 | NCE | 5.6397 | 48.8561 | x | | |
| France | FRAF198 | NCE | 5.6397 | 48.8561 | x | | |
| France | FRBe19 | NCE | 5.2656 | 45.2033 | x | | |
| France | FRBe20 | NCE | 5.2656 | 45.2033 | x | | |
| France | FRBe21 | NCE | 5.2656 | 45.2033 | x | | |
| France | FRBl100 | Pyr | -3.1958 | 47.3375 | x | B lle | Small |
| France | FRBl101 | Pyr | -3.1958 | 47.3375 | x | B lle | Small |
| France | FRBl102 | Pyr | -3.1958 | 47.3375 | x | B lle | Small |
| France | FRBl103 | Pyr | -3.1958 | 47.3375 | x | B lle | Small |
| France | FRBl104 | Pyr | -3.1958 | 47.3375 | x | B lle | Small |
| France | FRBl105 | Pyr | -3.1958 | 47.3375 | x | B lle | Small |
| France | FRBl106 | Pyr | -3.1958 | 47.3375 | x | B lle | Small |
| France | FRBl107 | Pyr | -3.1958 | 47.3375 | x | B lle | Small |
| France | FRBl108 | Pyr | -3.1958 | 47.3375 | x | B lle | Small |
| France | FRBl109 | Pyr | -3.1958 | 47.3375 | x | B lle | Small |
| France | FRBl110 | Pyr | -3.1958 | 47.3375 | x | B lle | Small |
| France | FRBl111 | Pyr | -3.1958 | 47.3375 | x | B lle | Small |
| France | FRBl93 | Pyr | -3.1958 | 47.3375 | x | B lle | Small |
| France | FRBl94 | Pyr | -3.1958 | 47.3375 | x | B lle | Small |
| France | FRBl95 | Pyr | -3.1958 | 47.3375 | x | B lle | Small |
| France | FRBl96 | Pyr | -3.1958 | 47.3375 | x | B lle | Small |
| France | FRBl97 | Pyr | -3.1958 | 47.3375 | x | B lle | Small |
| France | FRBl98 | Pyr | -3.1958 | 47.3375 | x | B lle | Small |
| France | FRBl99 | Pyr | -3.1958 | 47.3375 | x | B lle | Small |
| France | FRBo141 | NCE | 5.7622 | 48.7475 | x | | |
| France | FRBo142 | NCE | 5.7622 | 48.7475 | x | | |
| France | FRBo143 | NCE | 5.7622 | 48.7475 | x | | |
| France | FRBo144 | NCE | 5.7622 | 48.7475 | x | | |
| France | FRBo145 | NCE | 5.7622 | 48.7475 | x | | |
| France | FRBo146 | NCE | 5.7622 | 48.7475 | x | | |
| France | FRBo147 | NCE | 5.7622 | 48.7475 | x | | |
| France | FRBo148 | NCE | 5.7622 | 48.7475 | x | | |
| France | FRBo149 | NCE | 5.7622 | 48.7475 | x | | |
| France | FRBo150 | NCE | 5.7622 | 48.7475 | x | | |
| France | FRBo151 | NCE | 5.7622 | 48.7475 | x | | |
| France | FRBo152 | NCE | 5.7622 | 48.7475 | x | | |
| France | FRBo153 | NCE | 5.7622 | 48.7475 | x | | |
| France | FRBo154 | NCE | 5.7622 | 48.7475 | x | | |
| France | FRBo155 | NCE | 5.7622 | 48.7475 | x | | |
| France | FRBo156 | NCE | 5.7622 | 48.7475 | x | | |
| France | FRBo157 | NCE | 5.7622 | 48.7475 | x | | |
| France | FRBo158 | NCE | 5.7622 | 48.7475 | x | | |
| France | FRBo159 | NCE | 5.7622 | 48.7475 | x | | |
| France | FRBo160 | NCE | 5.7622 | 48.7475 | x | | |
| France | FRBo161 | NCE | 5.7622 | 48.7475 | x | | |
| France | FRBo162 | NCE | 5.7622 | 48.7475 | x | | |
| France | FRBo163 | NCE | 5.7622 | 48.7475 | x | | |
| France | FRBo164 | NCE | 5.7622 | 48.7475 | x | | |
| France | FRBo165 | NCE | 5.7622 | 48.7475 | x | | |
| France | FRBo166 | NCE | 5.7622 | 48.7475 | x | | |
| France | FRCh043A | NCE | 5.4317 | 45.4447 | x | | |

| France | FRCh38 | NCE | 5.4317 | 45.4447 | x | | | |
|--------|--------|-----|--------|---------|---|---|--------|------|
| France | FRCh39 | NCE | 5.4317 | 45.4447 | x | | | |
| France | FRCh40 | NCE | 5.4317 | 45.4447 | x | | | |
| France | FRCh41 | NCE | 5.4317 | 45.4447 | x | | | |
| France | FRCh42 | NCE | 5.4317 | 45.4447 | x | | | |
| France | FRCh44 | NCE | 5.4317 | 45.4447 | x | | | |
| France | FRCh45 | NCE | 5.4317 | 45.4447 | x | | | |
| France | FRCh46 | NCE | 5.4317 | 45.4447 | x | | | |
| France | FRCh48 | NCE | 5.4317 | 45.4447 | x | | | |
| France | FRCh49 | NCE | 5.4317 | 45.4447 | x | | | |
| France | FRCh50 | NCE | 5.4317 | 45.4447 | x | | | |
| France | FRCh51 | NCE | 5.4317 | 45.4447 | x | | | |
| France | FRCh52 | NCE | 5.4317 | 45.4447 | x | | | |
| France | FRCh54 | NCE | 5.4317 | 45.4447 | x | | | |
| France | FRCh55 | NCE | 5.4317 | 45.4447 | x | | | |
| France | FRCh56 | NCE | 5.4317 | 45.4447 | x | | | |
| France | FRCh58 | NCE | 5.4317 | 45.4447 | x | | | |
| France | FRCM1 | NCE | 4.7275 | 45.3847 | x | | | |
| France | FRCu136 | Pyr | 3.8850 | 44.9894 | x | | France | Cont |
| France | FRDi3003 | NCI | 6.1432 | 46.3568 | x | x | | |
| France | FRFS24 | NCE | 4.6928 | 45.2975 | x | | | |
| France | FRFS25 | NCE | 4.6928 | 45.2975 | x | | | |
| France | FRFS26 | NCE | 4.6928 | 45.2975 | x | | | |
| France | FRFT12 | NCE | 5.5244 | 45.5275 | x | | | |
| France | FRGL035A | NCE | 5.4206 | 45.3992 | x | | | |
| France | FRGL33 | NCE | 5.4206 | 45.3992 | x | | | |
| France | FRGL34 | NCE | 5.4206 | 45.3992 | x | | | |
| France | FRHa59 | NCI | 6.2847 | 47.8436 | x | | | |
| France | FRHa60 | NCI | 6.2847 | 47.8436 | x | | | |
| France | FRHa61 | NCI | 6.2847 | 47.8436 | x | | | |
| France | FRHa62 | NCI | 6.2847 | 47.8436 | x | | | |
| France | FRHa63 | NCI | 6.2847 | 47.8436 | x | | | |
| France | FRHa64 | NCI | 6.2847 | 47.8436 | x | | | |
| France | FRHa65 | NCI | 6.2847 | 47.8436 | x | | | |
| France | FRLa87 | NCI | 5.4108 | 48.9397 | x | | | |
| France | FRLa88 | NCI | 5.4108 | 48.9397 | x | | | |
| France | FRLa89 | NCI | 5.4108 | 48.9397 | x | | | |
| France | FRLe37 | NCE | 5.1142 | 45.3000 | x | | | |
| France | FRLG003A | NCI | 5.9031 | 45.0919 | x | | | |
| France | FRLG003B | NCI | 5.9031 | 45.0919 | x | | | |
| France | FRLG003C | NCI | 5.9031 | 45.0919 | x | | | |
| France | FRLo13 | NCE | 5.3483 | 45.4197 | x | | | |
| France | FRLo14 | NCE | 5.3483 | 45.4197 | x | | | |
| France | FRLV27 | NCI | 5.5891 | 45.1280 | x | | | |
| France | FRMu16 | NCE | 5.3158 | 45.2139 | x | | | |
| France | FRMu36 | NCE | 5.3158 | 45.2139 | x | | | |
| France | FROR15 | NCI | 5.8706 | 44.9219 | x | | | |
| France | FRPD3082 | NCE | 6.3818 | 49.0071 | x | x | | |
| France | FRSA079A | Pyr | 1.1675 | 44.8867 | x | | France | Cont |
| France | FRSA079B | Pyr | 1.1675 | 44.8867 | x | | France | Cont |
| France | FRSA168 | Pyr | 1.1675 | 44.8867 | x | | France | Cont |
| France | FRSA169 | Pyr | 1.1675 | 44.8867 | x | | France | Cont |
| France | FRSA171 | Pyr | 1.1675 | 44.8867 | x | | France | Cont |
| France | FRSA172 | Pyr | 1.1675 | 44.8867 | x | | France | Cont |
| France | FRSA173 | Pyr | 1.1675 | 44.8867 | x | | France | Cont |
| France | FRSA68 | Pyr | 1.1675 | 44.8867 | x | | France | Cont |
| France | FRSA69 | Pyr | 1.1675 | 44.8867 | x | | France | Cont |
| France | FRSA70 | Pyr | 1.1675 | 44.8867 | x | | France | Cont |
| France | FRSA71 | Pyr | 1.1675 | 44.8867 | x | | France | Cont |
| France | FRSA72 | Pyr | 1.1675 | 44.8867 | x | | France | Cont |

| France | FRSA74 | Pyr | 1.1675 | 44.8867 | x | | France | Cont |
|--------|--------|-----|--------|---------|---|---|--------|------|
| France | FRSA75 | Pyr | 1.1675 | 44.8867 | x | | France | Cont |
| France | FRSA77 | Pyr | 1.1675 | 44.8867 | x | | France | Cont |
| France | FRSA78 | Pyr | 1.1675 | 44.8867 | x | | France | Cont |
| France | FRSA81 | Pyr | 1.1675 | 44.8867 | x | | France | Cont |
| France | FRSA82 | Pyr | 1.1675 | 44.8867 | x | | France | Cont |
| France | FRSA83 | Pyr | 1.1675 | 44.8867 | x | | France | Cont |
| France | FRSA85 | Pyr | 1.1675 | 44.8867 | x | | France | Cont |
| France | FRSJ030A | NCE | 5.1386 | 45.5031 | x | | | |
| France | FRSJ030B | NCE | 5.1386 | 45.5031 | x | | | |
| France | FRSJ030C | NCE | 5.1386 | 45.5031 | x | | | |
| France | FRSJ10 | NCE | 5.1386 | 45.5031 | x | | | |
| France | FRSJ11 | NCE | 5.1386 | 45.5031 | x | | | |
| France | FRSJ28 | NCE | 5.1386 | 45.5031 | x | | | |
| France | FRSJ31 | NCE | 5.1386 | 45.5031 | x | | | |
| France | FRSJ32 | NCE | 5.1386 | 45.5031 | x | | | |
| France | FRSJ9 | NCE | 5.1386 | 45.5031 | x | | | |
| France | FRSN117 | Pyr | 3.7900 | 44.8911 | x | | France | Cont |
| France | FRSN118 | Pyr | 3.7900 | 44.8911 | x | | France | Cont |
| France | FRSN120 | Pyr | 3.7900 | 44.8911 | x | | France | Cont |
| France | FRSN121 | Pyr | 3.7900 | 44.8911 | x | | France | Cont |
| France | FRSN123 | Pyr | 3.7900 | 44.8911 | x | | France | Cont |
| France | FRSN124 | Pyr | 3.7900 | 44.8911 | x | | France | Cont |
| France | FRSN125 | Pyr | 3.7900 | 44.8911 | x | | France | Cont |
| France | FRSN126 | Pyr | 3.7900 | 44.8911 | x | | France | Cont |
| France | FRSN127 | Pyr | 3.7900 | 44.8911 | x | | France | Cont |
| France | FRSo091A | NCI | 7.3361 | 47.4839 | x | | | |
| France | FRSo091B | NCI | 7.3361 | 47.4839 | x | | | |
| France | FRSo90 | NCI | 7.3361 | 47.4839 | x | | | |
| France | FRSo92 | NCI | 7.3361 | 47.4839 | x | | | |
| France | FRSS3 | NCE | 4.2125 | 45.9486 | x | | | |
| France | FRSS4 | NCE | 4.2125 | 45.9486 | x | | | |
| France | FRSS5 | NCE | 4.2125 | 45.9486 | x | | | |
| France | FRSS6 | NCE | 4.2125 | 45.9486 | x | | | |
| France | FRVa17 | NCE | 5.4114 | 45.2569 | x | | | |
| France | FRVa18 | NCE | 5.4114 | 45.2569 | x | | | |
| France | FRVe129 | NCE | 4.6633 | 45.3692 | x | | | |
| France | FRVe22 | NCE | 4.6633 | 45.3692 | x | | | |
| France | FRVe23 | NCE | 4.6633 | 45.3692 | x | | | |
| France | FRVy7 | NCE | 4.6572 | 45.7383 | x | | | |
| Germany | DEEb3996 | NCE | 13.8109 | 52.8331 | x | x | | |
| Greece | GREp6406 | Balk | 21.1692 | 39.7705 | x | x | | |
| Ireland | IECL1 | Ire | -7.9480 | 53.2678 | x | x | Cloghan | Large |
| Ireland | IECL10 | Ire | -7.9480 | 53.2678 | x | x | Cloghan | Large |
| Ireland | IECL11 | Ire | -7.9480 | 53.2678 | x | x | Cloghan | Large |
| Ireland | IECL12 | Ire | -7.9480 | 53.2678 | x | x | Cloghan | Large |
| Ireland | IECL13 | Ire | -7.9480 | 53.2678 | x | x | Cloghan | Large |
| Ireland | IECL14 | Ire | -7.9480 | 53.2678 | x | x | Cloghan | Large |
| Ireland | IECL15 | Ire | -7.9480 | 53.2678 | x | x | Cloghan | Large |
| Ireland | IECL16 | Ire | -7.9480 | 53.2678 | x | x | Cloghan | Large |
| Ireland | IECL17 | Ire | -7.9480 | 53.2678 | x | x | Cloghan | Large |
| Ireland | IECL18 | Ire | -7.9480 | 53.2678 | x | x | Cloghan | Large |
| Ireland | IECL19 | Ire | -7.9480 | 53.2678 | x | x | Cloghan | Large |
| Ireland | IECL2 | Ire | -7.9480 | 53.2678 | x | x | Cloghan | Large |
| Ireland | IECL20 | Ire | -7.9480 | 53.2678 | x | x | Cloghan | Large |
| Ireland | IECL3 | Ire | -7.9480 | 53.2678 | x | x | Cloghan | Large |
| Ireland | IECL4 | Ire | -7.9480 | 53.2678 | x | x | Cloghan | Large |
| Ireland | IECL5 | Ire | -7.9480 | 53.2678 | x | x | Cloghan | Large |
| Ireland | IECL6 | Ire | -7.9480 | 53.2678 | x | x | Cloghan | Large |
| Ireland | IECL7 | Ire | -7.9480 | 53.2678 | x | x | Cloghan | Large |

| Ireland | IECL8 | Ire | -7.9480 | 53.2678 | x | x | Cloghan | Large |
|---------|-------|-----|---------|---------|---|---|---------|-------|
| Ireland | IECL9 | Ire | -7.9480 | 53.2678 | x | x | Cloghan | Large |
| Ireland | IEDY1 | Ire | -7.2500 | 55.0000 | x | x | Derry | Large |
| Ireland | IEDY10 | Ire | -7.2500 | 55.0000 | x | x | Derry | Large |
| Ireland | IEDY11 | Ire | -7.2500 | 55.0000 | x | x | Derry | Large |
| Ireland | IEDY12 | Ire | -7.2500 | 55.0000 | x | x | Derry | Large |
| Ireland | IEDY13 | Ire | -7.2500 | 55.0000 | x | x | Derry | Large |
| Ireland | IEDY14 | Ire | -7.2500 | 55.0000 | x | x | Derry | Large |
| Ireland | IEDY15 | Ire | -7.2500 | 55.0000 | x | x | Derry | Large |
| Ireland | IEDY16 | Ire | -7.2500 | 55.0000 | x | | Derry | Large |
| Ireland | IEDY17 | Ire | -7.2500 | 55.0000 | x | x | Derry | Large |
| Ireland | IEDY18 | Ire | -7.2500 | 55.0000 | x | x | Derry | Large |
| Ireland | IEDY19 | Ire | -7.2500 | 55.0000 | x | x | Derry | Large |
| Ireland | IEDY2 | Ire | -7.2500 | 55.0000 | x | x | Derry | Large |
| Ireland | IEDY20 | Ire | -7.2500 | 55.0000 | x | x | Derry | Large |
| Ireland | IEDY3 | Ire | -7.2500 | 55.0000 | x | x | Derry | Large |
| Ireland | IEDY4 | Ire | -7.2500 | 55.0000 | x | x | Derry | Large |
| Ireland | IEDY5 | Ire | -7.2500 | 55.0000 | x | x | Derry | Large |
| Ireland | IEDY6 | Ire | -7.2500 | 55.0000 | x | x | Derry | Large |
| Ireland | IEDY7 | Ire | -7.2500 | 55.0000 | x | x | Derry | Large |
| Ireland | IEDY8 | Ire | -7.2500 | 55.0000 | x | x | Derry | Large |
| Ireland | IEDY9 | Ire | -7.2500 | 55.0000 | x | | Derry | Large |
| Ireland | IEE1RV | Ire | -8.3505 | 54.9503 | x | x | Gweed | Large |
| Ireland | IEE2RV | Ire | -7.5152 | 54.9754 | x | x | Derry | Large |
| Ireland | IEE3RV | Ire | -7.5152 | 54.9754 | x | x | Derry | Large |
| Ireland | IEE4RV | Ire | -7.5152 | 54.9754 | x | x | Derry | Large |
| Ireland | IEGw1 | Ire | -8.2300 | 55.0500 | x | x | Gweed | Large |
| Ireland | IEGw17 | Ire | -8.2300 | 55.0500 | x | x | Gweed | Large |
| Ireland | IEGw1RV | Ire | -8.3833 | 55.0500 | x | x | Gweed | Large |
| Ireland | IEGw26 | Ire | -8.2300 | 55.0500 | x | x | Gweed | Large |
| Ireland | IEGw3 | Ire | -8.2300 | 55.0500 | x | x | Gweed | Large |
| Ireland | IEGw4 | Ire | -8.2300 | 55.0500 | x | x | Gweed | Large |
| Ireland | IEGw43 | Ire | -8.2300 | 55.0500 | x | x | Gweed | Large |
| Ireland | IEGw46 | Ire | -8.2300 | 55.0500 | x | x | Gweed | Large |
| Ireland | IEGw5 | Ire | -8.2300 | 55.0500 | x | x | Gweed | Large |
| Ireland | IEGw51 | Ire | -8.2300 | 55.0500 | x | x | Gweed | Large |
| Ireland | IEGw51A | Ire | -8.2300 | 55.0500 | x | x | Gweed | Large |
| Ireland | IEGw51B | Ire | -8.2300 | 55.0500 | x | x | Gweed | Large |
| Ireland | IEGw51C | Ire | -8.2300 | 55.0500 | x | x | Gweed | Large |
| Ireland | IEGw55 | Ire | -8.2300 | 55.0500 | x | x | Gweed | Large |
| Ireland | IEGw55B | Ire | -8.2300 | 55.0500 | x | x | Gweed | Large |
| Ireland | IEGw5b | Ire | -8.2300 | 55.0500 | x | | Gweed | Large |
| Ireland | IEGw5C | Ire | -8.2300 | 55.0500 | x | x | Gweed | Large |
| Ireland | IEGw64 | Ire | -8.2300 | 55.0500 | x | x | Gweed | Large |
| Ireland | IEGwTILES | Ire | -8.2300 | 55.0500 | x | x | Gweed | Large |
| Italy | ITAn23 | NCI | 7.6966 | 45.8224 | x | | | |
| Italy | ITCh16 | NCI | 7.6229 | 45.6217 | x | | | |
| Italy | ITCh17 | NCI | 7.6229 | 45.6217 | x | x | | |
| Italy | ITCh18 | NCI | 7.6229 | 45.6217 | x | | | |
| Italy | ITGa33 | NCI | 7.8482 | 45.8519 | | x | | |
| Italy | ITGa36 | NCI | 7.8482 | 45.8519 | x | x | | |
| Italy | ITGa38 | NCI | 7.8482 | 45.8519 | x | x | | |
| Italy | ITMa9815 | NCI | 14.1157 | 42.0834 | x | x | | |
| Italy | ITMa9830 | NCI | 14.1157 | 42.0834 | x | x | | |
| Italy | ITMa9832 | NCI | 14.1157 | 42.0834 | x | x | | |
| Italy | ITMa9833 | NCI | 14.1157 | 42.0834 | x | x | | |
| Italy | ITMa9834 | NCI | 14.1157 | 42.0834 | x | x | | |
| Italy | ITMa9835 | NCI | 14.1157 | 42.0834 | x | x | | |
| Italy | ITMa9849 | NCI | 14.1157 | 42.0834 | x | x | | |
| Italy | ITMa9850 | NCI | 14.1157 | 42.0834 | x | x | | |

| Italy | ITMC32500 | NCI | 10.8360 | 46.2387 | x | |
|-------|-----------|-----|---------|---------|---|---|
| Italy | ITPr0001 | NCI | 10.2480 | 46.2870 | x | x |
| Italy | ITPr0004 | NCE | 10.2480 | 46.2870 | x | x |
| Italy | ITSC54303 | NCI | 8.8320 | 44.4773 | x | x |
| Italy | ITSi11 | SIta | 16.4911 | 39.3522 | x | x |
| Italy | ITSi17 | SIta | 16.4911 | 39.3522 | x | x |
| Italy | ITSi21 | SIta | 16.4911 | 39.3522 | x | x |
| Italy | ITTg47 | NCI | 7.5712 | 45.8072 | x | x |
| Italy | ITTg48 | NCI | 7.5712 | 45.8072 | x | x |
| Italy | ITTg49 | NCE | 7.5712 | 45.8072 | x | x |
| Italy | ITTr17692 | NCI | 11.8333 | 46.2500 | x | |
| Italy | ITVB54317 | NCI | 9.0648 | 44.5553 | x | x |
| Macedonia | MKBi2450 | Balk | 20.7688 | 41.5169 | x | |
| Macedonia | MKJa9212 | Balk | 21.4189 | 41.6891 | x | x |
| Macedonia | MKJa9222 | Balk | 21.4189 | 41.6891 | x | |
| Macedonia | MKJa9223 | Balk | 21.4189 | 41.6891 | x | x |
| Macedonia | MKKo194 | Balk | 22.3942 | 41.1544 | x | x |
| Macedonia | MKKo195 | Balk | 22.3942 | 41.1544 | x | x |
| Macedonia | MKPe3834 | Balk | 21.1675 | 41.0089 | x | |
| Macedonia | MKPe3835 | Balk | 21.1675 | 41.0089 | x | x |
| Macedonia | MKPe3836 | Balk | 21.1675 | 41.0089 | | x |
| Macedonia | MKPe3896 | Balk | 21.1675 | 41.0089 | x | x |
| Macedonia | MKPe9494 | Balk | 21.1675 | 41.0089 | x | x |
| Macedonia | MKPe9505 | Balk | 21.1675 | 41.0089 | x | x |
| Macedonia | MKPe9645 | Balk | 21.1675 | 41.0089 | x | x |
| Montenegro | MEBj381 | Balk | 19.7011 | 42.8657 | x | x |
| Montenegro | MEBj382 | Balk | 19.7011 | 42.8657 | x | x |
| Montenegro | MEBj383 | Balk | 19.7011 | 42.8657 | x | x |
| Montenegro | MEDu3403 | Balk | 19.0412 | 43.1455 | x | x |
| Montenegro | MEDu3430 | Balk | 19.0412 | 43.1455 | x | x |
| Poland | POBie115 | NCE | 22.5734 | 53.6439 | x | x |
| Poland | POBie140 | NCE | 22.5734 | 53.6439 | x | |
| Poland | POBie141 | NCE | 22.5734 | 53.6439 | x | x |
| Poland | POBie151 | NCE | 22.5734 | 53.6439 | x | x |
| Poland | POBie355 | NCE | 22.5734 | 53.6439 | x | x |
| Poland | POBie373 | NCE | 22.5734 | 53.6439 | x | x |
| Poland | POBie376 | NCE | 22.5734 | 53.6439 | x | x |
| Poland | POBie377 | NCE | 22.5734 | 53.6439 | x | x |
| Poland | POBie392 | NCE | 22.5734 | 53.6439 | x | x |
| Poland | POBie912 | NCE | 22.5734 | 53.6439 | x | x |
| Poland | POBPN315 | NCE | 23.9001 | 52.7095 | x | x |
| Poland | POBPN316 | NCE | 23.9001 | 52.7095 | x | x |
| Poland | POBPN318 | NCE | 23.9001 | 52.7095 | x | x |
| Poland | POBPN319 | NCE | 23.9001 | 52.7095 | x | x |
| Poland | POBPN320 | NCE | 23.9001 | 52.7095 | x | x |
| Poland | POBPN345 | NCE | 23.9001 | 52.7095 | x | x |
| Poland | POBPN346 | NCE | 23.9001 | 52.7095 | x | x |
| Poland | POBPN347 | NCE | 23.9001 | 52.7095 | x | x |
| Poland | POBPN402 | NCE | 23.9001 | 52.7095 | x | x |
| Poland | POBPN411 | NCE | 23.9001 | 52.7095 | x | x |
| Serbia | RSBe178 | Balk | 20.0800 | 45.6147 | x | x |
| Serbia | RSKo40169 | Balk | 20.9981 | 44.7290 | x | |
| Serbia | RSMF53266 | Balk | 19.6625 | 45.1707 | x | x |
| Serbia | RSMF566 | Balk | 19.6625 | 45.1707 | x | x |
| Serbia | RSMK10066 | Balk | 20.8106 | 43.2967 | x | x |
| Serbia | RSMK1078 | Balk | 20.8106 | 43.2967 | x | x |
| Serbia | RSMK1276 | Balk | 20.8106 | 43.2967 | x | x |
| Serbia | RSMK17377 | Balk | 20.8106 | 43.2967 | x | x |
| Serbia | RSMK2449 | Balk | 20.8106 | 43.2967 | x | x |
| Serbia | RSMK578 | Balk | 20.8106 | 43.2967 | x | x |

| Serbia | RSMK678 | Balk | 20.8106 | 43.2967 | x | x | | |
| Serbia | RSMP35866 | Balk | 20.3619 | 42.8397 | x | | | |
| Serbia | RSMP35966 | Balk | 20.3619 | 42.8397 | x | x | | |
| Serbia | RSVa7841 | Balk | 19.7363 | 44.3148 | x | x | | |
| Serbia | RSVa7842 | Balk | 19.7363 | 44.3148 | x | x | | |
| Serbia | RSVa7855 | Balk | 19.7363 | 44.3148 | x | x | | |
| Slovenia | SICe142 | NCl | 14.9457 | 46.1730 | | x | | |
| Slovenia | SIDo2040 | NCl | 14.7974 | 45.5015 | x | x | | |
| Slovenia | SIDo2041 | NCl | 14.7974 | 45.5015 | | x | | |
| Slovenia | SIDr2778 | NCl | 14.0292 | 46.3588 | x | x | | |
| Slovenia | SIDr2779 | NCl | 14.0292 | 46.3588 | x | x | | |
| Slovenia | SIGo2042 | Balk | 15.5612 | 45.8578 | x | x | | |
| Slovenia | SIGr973 | NCl | 14.1325 | 46.1055 | x | x | | |
| Slovenia | SIGr974 | NCl | 14.1325 | 46.1055 | x | x | | |
| Slovenia | SIHo15910 | Balk | 16.3295 | 46.8110 | x | x | | |
| Slovenia | SIIg1563 | NCl | 14.5429 | 45.9470 | x | x | | |
| Slovenia | SIIg1564 | NCl | 14.5429 | 45.9470 | x | | | |
| Slovenia | SIIg1565 | NCl | 14.5429 | 45.9470 | x | x | | |
| Slovenia | SIIg1628 | NCl | 14.5429 | 45.9470 | x | x | | |
| Slovenia | SIIg1648 | NCl | 14.5429 | 45.9470 | x | x | | |
| Slovenia | SIIg1766 | NCl | 14.5429 | 45.9470 | x | x | | |
| Slovenia | SIIg1773 | NCl | 14.5429 | 45.9470 | x | x | | |
| Slovenia | SIIg1847 | NCl | 14.5429 | 45.9470 | x | x | | |
| Slovenia | SIIg2143 | NCl | 14.5429 | 45.9470 | x | x | | |
| Slovenia | SIKn6167 | NCl | 13.7844 | 46.4695 | x | x | | |
| Slovenia | SIKn6380 | NCl | 13.7844 | 46.4695 | x | x | | |
| Slovenia | SIKo6781 | NCl | 14.8203 | 45.6570 | x | | | |
| Slovenia | SIKo6782 | NCl | 14.8203 | 45.6570 | x | x | | |
| Slovenia | SIKo6827 | NCl | 14.8203 | 45.6570 | x | x | | |
| Slovenia | SIKo6845 | NCl | 14.8203 | 45.6570 | x | x | | |
| Slovenia | SIKr1042 | NCl | 14.2471 | 45.8215 | x | x | | |
| Slovenia | SIKr14709 | Balk | 15.4761 | 45.8956 | x | x | | |
| Slovenia | SILe1145 | Balk | 16.4575 | 46.5511 | x | x | | |
| Slovenia | SILe1146 | Balk | 16.4575 | 46.5511 | x | x | | |
| Slovenia | SILe1147 | Balk | 16.4575 | 46.5511 | x | x | | |
| Slovenia | SILe1163 | Balk | 16.4575 | 46.5511 | x | x | | |
| Slovenia | SING54318 | NCl | 13.6571 | 45.9447 | x | x | | |
| Slovenia | SING54320 | NCl | 13.6571 | 45.9447 | x | x | | |
| Slovenia | SIPd1045 | NCl | 13.9402 | 45.5129 | x | x | | |
| Slovenia | SIPd1372 | NCl | 13.9402 | 45.5129 | x | x | | |
| Slovenia | SIPd1374 | NCl | 13.9402 | 45.5129 | x | x | | |
| Slovenia | SIPh3126 | Balk | 15.2567 | 46.5193 | x | x | | |
| Slovenia | SIPh3131 | Balk | 15.2567 | 46.5193 | x | x | | |
| Slovenia | SIPh3489 | Balk | 15.2567 | 46.5193 | x | x | | |
| Slovenia | SIPk16394 | NCl | 14.0292 | 46.3588 | x | x | | |
| Slovenia | SIRa16104 | Balk | 15.3403 | 45.6859 | x | x | | |
| Slovenia | SISe16100 | Balk | 15.2350 | 45.5078 | x | | | |
| Slovenia | SISe16101 | Balk | 15.2350 | 45.5078 | x | x | | |
| Slovenia | SISl1378 | NCl | 13.9402 | 45.5129 | x | x | | |
| Slovenia | SISl1380 | NCl | 13.9402 | 45.5129 | x | | | |
| Slovenia | SISl1381 | NCl | 13.9402 | 45.5129 | x | x | | |
| Slovenia | SISl1382 | NCl | 13.9402 | 45.5129 | | x | | |
| Slovenia | SISl1383 | NCl | 13.9402 | 45.5129 | x | x | | |
| Slovenia | SISl1384 | NCl | 13.9402 | 45.5129 | | x | | |
| Slovenia | SISn1043 | NCl | 14.4014 | 45.5733 | x | x | | |
| Slovenia | SISn1044 | NCl | 14.4014 | 45.5733 | x | x | | |
| Spain | ESArE135 | Iber | -4.2000 | 43.0333 | x | | | |
| Spain | ESCoE138 | Pyr | -3.6272 | 43.0193 | x | | Nav | Cont |
| Spain | ESEM69 | Pyr | -3.4503 | 43.1423 | x | x | Nav | Cont |
| Spain | ESNa1131 | Pyr | -1.6455 | 43.1757 | x | x | Nav | Cont |

| Spain | ESNa1286 | Pyr | -1.6455 | 43.1757 | x | x | Nav | Cont |
|---|---|---|---|---|---|---|---|---|
| Spain | ESNa137 | Pyr | -1.6455 | 43.1757 | x | x | Nav | Cont |
| Spain | ESNa1379 | Pyr | -1.6455 | 43.1757 | x | x | Nav | Cont |
| Spain | ESNa1576 | Pyr | -1.6455 | 43.1757 | x | x | Nav | Cont |
| Spain | ESNa1577 | Pyr | -1.6455 | 43.1757 | x | x | Nav | Cont |
| Spain | ESNa1579 | Pyr | -1.6455 | 43.1757 | x | x | Nav | Cont |
| Spain | ESNa172 | Pyr | -1.6455 | 43.1757 | x | x | Nav | Cont |
| Spain | ESNa1757 | Pyr | -1.6455 | 43.1757 | x | x | Nav | Cont |
| Spain | ESNa1758 | Pyr | -1.6455 | 43.1757 | x | x | Nav | Cont |
| Spain | ESNa239 | Pyr | -1.6455 | 43.1757 | | x | Nav | Cont |
| Spain | ESNa240 | Pyr | -1.6455 | 43.1757 | x | x | Nav | Cont |
| Spain | ESNa318 | Pyr | -1.6455 | 43.1757 | x | x | Nav | Cont |
| Spain | ESNa399 | Pyr | -1.6455 | 43.1757 | x | x | Nav | Cont |
| Spain | ESNa406 | Pyr | -1.6455 | 43.1757 | x | x | Nav | Cont |
| Spain | ESNa460 | Pyr | -1.6455 | 43.1757 | x | x | Nav | Cont |
| Spain | ESNa461 | Pyr | -1.6455 | 43.1757 | x | x | Nav | Cont |
| Spain | ESNa463 | Pyr | -1.6455 | 43.1757 | x | x | Nav | Cont |
| Spain | ESNa47 | Pyr | -1.6455 | 43.1757 | x | x | Nav | Cont |
| Spain | ESNa509 | Pyr | -1.6455 | 43.1757 | x | x | Nav | Cont |
| Spain | ESNa598 | Pyr | -1.6455 | 43.1757 | x | x | Nav | Cont |
| Spain | ESNa633 | Pyr | -1.6455 | 43.1757 | x | x | Nav | Cont |
| Spain | ESNa739 | Pyr | -1.6455 | 43.1757 | x | x | Nav | Cont |
| Spain | ESNa752 | Pyr | -1.6455 | 43.1757 | x | x | Nav | Cont |
| Spain | ESNa798 | Pyr | -1.6455 | 43.1757 | x | x | Nav | Cont |
| Spain | ESNa803 | Pyr | -1.6455 | 43.1757 | x | x | Nav | Cont |
| Spain | ESNa861 | Pyr | -1.6455 | 43.1757 | x | x | Nav | Cont |
| Spain | ESPE47 | Iber | -4.9997 | 43.1049 | x | x | | |
| Spain | ESPE57 | Iber | -4.9997 | 43.1049 | x | x | | |
| Spain | ESRa0640 | Iber | -3.8794 | 40.9036 | x | x | | |
| Spain | ESRa2653 | Iber | -3.8794 | 40.9036 | x | x | | |
| Spain | ESRa3443 | Iber | -3.8794 | 40.9036 | x | x | | |
| Spain | ESRa3444 | Iber | -3.8794 | 40.9036 | x | x | | |
| Spain | ESRa3445 | Iber | -3.8794 | 40.9036 | x | x | | |
| Spain | ESRa3446 | Iber | -3.8794 | 40.9036 | x | x | | |
| Spain | ESRa3447 | Iber | -3.8794 | 40.9036 | x | x | | |
| Spain | ESRa3448 | Iber | -3.8794 | 40.9036 | x | x | | |
| Spain | ESRa3449 | Iber | -3.8794 | 40.9036 | x | x | | |
| Spain | ESRa3451 | Iber | -3.8794 | 40.9036 | x | x | | |
| Switzerland | CHBa0441 | NCE | 6.2311 | 46.4628 | x | x | | |
| Switzerland | CHBa1816 | NCE | 6.2311 | 46.4628 | x | x | | |
| Switzerland | CHBa1817 | NCE | 6.2311 | 46.4628 | x | x | | |
| Switzerland | CHBa1818 | NCE | 6.2311 | 46.4628 | x | x | | |
| Switzerland | CHBa1819 | NCE | 6.2311 | 46.4628 | x | x | | |
| Switzerland | CHBa1820 | NCE | 6.2311 | 46.4628 | x | x | | |
| Switzerland | CHBa1821 | NCE | 6.2311 | 46.4628 | x | x | | |
| Switzerland | CHBa3002 | NCE | 6.2311 | 46.4628 | x | x | | |
| Switzerland | CHCh7622 | NCE | 6.9974 | 46.9327 | x | x | | |
| Switzerland | CHPo7628 | NCE | 6.9974 | 46.9327 | x | x | | |
| Switzerland | CHVI4746 | NCI | 6.8927 | 46.2043 | x | x | | |
| Switzerland | CHVI4747 | NCE | 6.8927 | 46.2043 | x | x | | |
| Switzerland | CHVI4748 | NCI | 6.8927 | 46.2043 | x | x | | |
| Turkey | Svolnu290 | Out | 31.7238 | 40.8988 | x | x | | |
| Turkey | Svolnu302 | Out | 31.7238 | 40.8988 | x | x | | |
| Turkey | Svolnu303 | Out | 31.7238 | 40.8988 | x | x | | |
| Turkey | Svolnu304 | Out | 31.7238 | 40.8988 | x | x | | |
| Turkey | Svolnu312 | Out | 31.7238 | 40.8988 | x | x | | |
| Turkey | Svolnu313 | Out | 31.7238 | 40.8988 | x | x | | |
| Turkey | Svolnu392 | Out | 31.7238 | 40.8988 | x | x | | |
| Turkey | Svolnu393 | Out | 31.7238 | 40.8988 | x | x | | |

NCE = North-Central European

| | |
|---|---|
| NCI = North-Central Italian | |
| SIta = South Italian | |
| Balk = Balkan | |
| Iber = Iberian | |
| Pyr = Pyrenean | |
| OM = Orkney Mainland | |
| OW = Orkney Westray | |
| OS = Orkney South Ronaldsay | |
| Grimn = Grimness | |
| Wind = Windwick | |
| Main1 = Mainland1 | |
| Main2 = Mainland2 | |
| Swart = Swartmill | |
| Gweed = Gweedore | |
| Nav = Navarra | |
| B Île = Belle Île | |
| Cont = Continental | |
| Large = Large island size | |
| Med = Medium island size | |
| Small = Small island size | |

X = sample available

**Appendix 5.** Analysis of Variance (ANOVA) post-hoc tests on Centroid Size (CS) for *Sorex minutus* mandibles.

Parametric ANOVA post-hoc test (Tukey–Kramer) on CS for *S. minutus* mandibles (pairwise *p* values).

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 1-Balkan | - | | | | | | | | | | |
| 2-Iberian | 0.9996 | - | | | | | | | | | |
| 3-Ireland | 0.0000 | 0.0000 | - | | | | | | | | |
| 4-North-Central European | 0.0000 | 0.0000 | 0.1160 | - | | | | | | | |
| 5-North-Central Italian | 0.0000 | 0.4774 | 0.0000 | 0.0000 | - | | | | | | |
| 6-Orkney Mainland | 0.0000 | 0.0002 | 0.8594 | 0.0002 | 0.0006 | - | | | | | |
| 7-Orkney South Ronaldsay | 0.0000 | 0.0000 | 0.8693 | 0.9999 | 0.0000 | 0.0875 | - | | | | |
| 8-Orkney Westray | 0.0000 | 0.0006 | 0.7900 | 0.0004 | 0.0075 | 1.0000 | 0.0766 | - | | | |
| 9-Pyrenean | 0.0000 | 0.0545 | 0.0001 | 0.0000 | 0.7729 | 0.1705 | 0.0000 | 0.4400 | - | | |
| 10-South Italian | 1.0000 | 1.0000 | 0.0378 | 0.0021 | 0.9282 | 0.1658 | 0.0066 | 0.2121 | 0.6731 | - | |
| 11-Outgroup | 0.9981 | 0.9691 | 0.0000 | 0.0000 | 0.0318 | 0.0000 | 0.0000 | 0.0000 | 0.0019 | 1.0000 | - |

Non-parametric ANOVA post-hoc test (Tamhane) on CS for *S. minutus* mandibles (pairwise *p* values).

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 1-Balkan | - | | | | | | | | | | |
| 2-Iberian | 1.0000 | - | | | | | | | | | |
| 3-Ireland | 0.0000 | 0.0000 | - | | | | | | | | |
| 4-North-Central European | 0.0000 | 0.0000 | 0.0421 | - | | | | | | | |
| 5-North-Central Italian | 0.0021 | 0.5889 | 0.0000 | 0.0000 | - | | | | | | |
| 6-Orkney Mainland | 0.0000 | 0.0004 | 0.9031 | 0.0000 | 0.0032 | - | | | | | |
| 7-Orkney South Ronaldsay | 0.0000 | 0.0000 | 0.4536 | 1.0000 | 0.0000 | 0.0005 | - | | | | |
| 8-Orkney Westray | 0.0000 | 0.0006 | 0.6080 | 0.0000 | 0.0081 | 1.0000 | 0.0001 | - | | | |
| 9-Pyrenean | 0.0000 | 0.0467 | 0.0004 | 0.0000 | 0.9999 | 0.2793 | 0.0000 | 0.5037 | - | | |
| 10-South Italian | 1.0000 | 1.0000 | 0.9840 | 0.9429 | 1.0000 | 0.9966 | 0.9569 | 0.9976 | 1.0000 | - | |
| 11-Outgroup | 1.0000 | 0.9975 | 0.0007 | 0.0003 | 0.0681 | 0.0019 | 0.0004 | 0.0025 | 0.0141 | 1.0000 | - |

**Appendix 6.** Multivariate Analysis of Variance (MANOVA) post-hoc tests on shape variables for *Sorex minutus* mandibles.

Parametric MANOVA post-hoc tests (Hotelling $T^2$) on shape variables among groups of *S. minutus* mandibles (pairwise *p* values).

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 1-Balkan | - | 0.0022 | 0.0000 | 0.0000 | 0.0002 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0964 | 0.0010 |
| 2-Iberian | 0.1217 | - | 0.0000 | 0.0002 | 0.0194 | 0.0000 | 0.0000 | 0.0000 | 0.0028 | Fail | Fail |
| 3-Ireland | 0.0000 | 0.0000 | - | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0005 | 0.0001 |
| 4-North-Central European | 0.0000 | 0.0098 | 0.0000 | - | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 |
| 5-North-Central Italian | 0.0124 | 1.0000 | 0.0000 | 0.0000 | - | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0556 | 0.0001 |
| 6-Orkney Mainland | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | - | 0.0000 | 0.0000 | 0.0000 | 0.1151 | 0.0003 |
| 7-Orkney South Ronaldsay | 0.0000 | 0.0026 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | - | 0.0000 | 0.0000 | 0.0360 | 0.0082 |
| 8-Orkney Westray | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | - | 0.0000 | 0.0044 | 0.0000 |
| 9-Pyrenean | 0.0000 | 0.1531 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | - | 0.0009 | 0.0089 |
| 10-South Italian | 1.0000 | Fail | 0.0293 | 0.0009 | 1.0000 | 1.0000 | 1.0000 | 0.2403 | 0.0503 | - | Fail |
| 11-Outgroup | 0.0534 | Fail | 0.0034 | 0.0000 | 0.0060 | 0.0149 | 0.4514 | 0.0000 | 0.4882 | Fail | - |

Below diagonal, Bonferroni corrected *p* values. Above diagonal, uncorrected *p* values.

Non-parametric MANOVA post-hoc tests on shape variables among groups of *S. minutus* mandibles (pairwise *p* values).

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 1-Balkan | - | 0.0158 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0049 | 0.0001 |
| 2-Iberian | 0.8690 | - | 0.0017 | 0.0064 | 0.1099 | 0.0007 | 0.0000 | 0.0000 | 0.0446 | 0.0191 | 0.0088 |
| 3-Ireland | 0.0000 | 0.0935 | - | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0006 | 0.0002 |
| 4-North-Central European | 0.0000 | 0.3520 | 0.0000 | - | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0024 | 0.0000 |
| 5-North-Central Italian | 0.0000 | 1.0000 | 0.0000 | 0.0000 | - | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0035 | 0.0001 |
| 6-Orkney Mainland | 0.0000 | 0.0385 | 0.0000 | 0.0000 | 0.0000 | - | 0.0006 | 0.0000 | 0.0000 | 0.0350 | 0.0052 |
| 7-Orkney South Ronaldsay | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0330 | - | 0.0000 | 0.0000 | 0.0005 | 0.0008 |
| 8-Orkney Westray | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | - | 0.0000 | 0.0001 | 0.0000 |
| 9-Pyrenean | 0.0000 | 1.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | - | 0.1044 | 0.0344 |
| 10-South Italian | 0.2695 | 1.0000 | 0.0330 | 0.1320 | 0.1925 | 1.0000 | 0.0275 | 0.0055 | 1.0000 | - | 0.5720 |
| 11-Outgroup | 0.0055 | 0.4840 | 0.0110 | 0.0000 | 0.0055 | 0.2860 | 0.0440 | 0.0000 | 1.0000 | 1.0000 | - |

Below diagonal, Bonferroni corrected *p* values. Above diagonal, uncorrected *p* values.

**Appendix 7.** Analysis of Variance (ANOVA) post-hoc tests on Centroid Size (CS) for *Sorex minutus* skulls.

Parametric ANOVA post-hoc test (Tukey–Kramer) on CS for *S. minutus* skulls (pairwise *p* values).

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 1-Balkan | - | | | | | | | | | | |
| 2-Iberian | 0.0009 | - | | | | | | | | | |
| 3-Ireland | 1.0000 | 0.0001 | - | | | | | | | | |
| 4-North-Central European | 0.0002 | 0.0000 | 0.0006 | - | | | | | | | |
| 5-North-Central Italian | 1.0000 | 0.0019 | 0.9921 | 0.0000 | - | | | | | | |
| 6-Orkney Mainland | 0.9890 | 0.0119 | 0.7933 | 0.0000 | 0.9997 | - | | | | | |
| 7-Orkney South Ronaldsay | 0.9946 | 0.0174 | 0.8760 | 0.0000 | 0.9999 | 1.0000 | - | | | | |
| 8-Orkney Westray | 0.5078 | 0.1505 | 0.1398 | 0.0000 | 0.7354 | 0.9860 | 0.9911 | - | | | |
| 9-Pyrenean | 1.0000 | 0.0060 | 0.9992 | 0.0006 | 1.0000 | 0.9999 | 1.0000 | 0.8741 | - | | |
| 10-South Italian | 1.0000 | 0.5180 | 1.0000 | 0.8526 | 1.0000 | 1.0000 | 1.0000 | 0.9994 | 1.0000 | - | |
| 11-Outgroup | 1.0000 | 0.1168 | 1.0000 | 0.1733 | 1.0000 | 1.0000 | 1.0000 | 0.9940 | 1.0000 | 1.0000 | - |

Non-parametric ANOVA post-hoc test (Tamhane) on CS for *S. minutus* skulls (pairwise *p* values).

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 1-Balkan | - | | | | | | | | | | |
| 2-Iberian | 0.3315 | - | | | | | | | | | |
| 3-Ireland | 1.0000 | 0.2030 | - | | | | | | | | |
| 4-North-Central European | 0.0594 | 0.0050 | 0.0556 | - | | | | | | | |
| 5-North-Central Italian | 1.0000 | 0.4253 | 1.0000 | 0.0082 | - | | | | | | |
| 6-Orkney Mainland | 1.0000 | 0.6699 | 0.9807 | 0.0037 | 1.0000 | - | | | | | |
| 7-Orkney South Ronaldsay | 1.0000 | 0.6218 | 0.3023 | 0.0007 | 1.0000 | 1.0000 | - | | | | |
| 8-Orkney Westray | 0.5362 | 0.9434 | 0.0003 | 0.0000 | 0.4399 | 1.0000 | 0.9116 | - | | | |
| 9-Pyrenean | 1.0000 | 0.4161 | 1.0000 | 0.0133 | 1.0000 | 1.0000 | 1.0000 | 0.5983 | - | | |
| 10-South Italian | 1.0000 | 0.9453 | 1.0000 | 0.9992 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | - | |
| 11-Outgroup | 1.0000 | 0.6249 | 1.0000 | 0.4519 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | - |

**Appendix 8.** Multivariate Analysis of Variance (MANOVA) post-hoc tests on shape variables for *Sorex minutus* skulls.

Parametric MANOVA post-hoc tests (Hotelling $T^2$) on shape variables among groups of *S. minutus* skulls (pairwise *p* values).

|  | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 1-Balkan | - | 0.0046 | 0.0000 | 0.0000 | 0.4654 | 0.0000 | 0.0000 | 0.0000 | 0.0003 | 0.1610 | 0.0233 |
| 2-Iberian | 0.2534 | - | 0.0002 | 0.0035 | 0.0011 | 0.0000 | 0.0000 | 0.0000 | 0.2338 | Fail | Fail |
| 3-Ireland | 0.0000 | 0.0120 | - | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0101 | 0.0000 |
| 4-North-Central European | 0.0000 | 0.1944 | 0.0000 | - | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.2786 | 0.0015 |
| 5-North-Central Italian | 1.0000 | 0.0607 | 0.0000 | 0.0000 | - | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0226 | 0.0078 |
| 6-Orkney Mainland | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | - | 0.0000 | 0.0000 | 0.0000 | 0.0035 | 0.0000 |
| 7-Orkney South Ronaldsay | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | - | 0.0000 | 0.0000 | 0.0764 | 0.0004 |
| 8-Orkney Westray | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | - | 0.0000 | 0.0097 | 0.0000 |
| 9-Pyrenean | 0.0157 | 1.0000 | 0.0000 | 0.0000 | 0.0010 | 0.0000 | 0.0000 | 0.0000 | - | Fail | 0.9437 |
| 10-South Italian | 1.0000 | Fail | 0.5540 | 1.0000 | 1.0000 | 0.1949 | 1.0000 | 0.5339 | Fail | - | Fail |
| 11-Outgroup | 1.0000 | Fail | 0.0000 | 0.0842 | 0.4299 | 0.0001 | 0.0193 | 0.0015 | 1.0000 | Fail | - |

Below diagonal, Bonferroni corrected *p* values. Above diagonal, uncorrected *p* values.

Non-parametric MANOVA post-hoc tests on shape variables among groups of *S. minutus* skulls (pairwise *p* values).

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 1-Balkan | - | 0.0002 | 0.0000 | 0.0000 | 0.0830 | 0.0000 | 0.0000 | 0.0000 | 0.0001 | 0.2640 | 0.0218 |
| 2-Iberian | 0.0110 | - | 0.0000 | 0.0000 | 0.0245 | 0.0000 | 0.0000 | 0.0000 | 0.0036 | 0.0324 | 0.0000 |
| 3-Ireland | 0.0000 | 0.0000 | - | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0030 | 0.0000 |
| 4-North-Central European | 0.0000 | 0.0000 | 0.0000 | - | 0.0001 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0070 | 0.0000 |
| 5-North-Central Italian | 1.0000 | 1.0000 | 0.0000 | 0.0055 | - | 0.0000 | 0.0000 | 0.0000 | 0.0008 | 0.1390 | 0.0108 |
| 6-Orkney Mainland | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | - | 0.0000 | 0.0000 | 0.0000 | 0.0004 | 0.0000 |
| 7-Orkney South Ronaldsay | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | - | 0.0000 | 0.0000 | 0.0006 | 0.0000 |
| 8-Orkney Westray | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | - | 0.0000 | 0.0000 | 0.0000 |
| 9-Pyrenean | 0.0055 | 0.1980 | 0.0000 | 0.0000 | 0.0440 | 0.0000 | 0.0000 | 0.0000 | - | 0.0762 | 0.0001 |
| 10-South Italian | 1.0000 | 1.0000 | 0.1650 | 0.3850 | 1.0000 | 0.0220 | 0.0330 | 0.0000 | 1.0000 | - | 0.9104 |
| 11-Outgroup | 1.0000 | 0.0000 | 0.0000 | 0.0000 | 0.5940 | 0.0000 | 0.0000 | 0.0000 | 0.0055 | 1.0000 | - |

Below diagonal, Bonferroni corrected *p* values. Above diagonal, uncorrected *p* values.

**Appendix 9.** Analysis of Variance (ANOVA) of *Sorex minutus* mandibles.

Parametric ANOVA post-hoc test (Tukey–Kramer) on CS for *S. minutus* mandibles (pairwise *p* values).

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1-Ireland Cloghan | - | | | | | | | | | | | |
| 2-Ireland Derry | 0.3242 | - | | | | | | | | | | |
| 3-Ireland Gweedore | 0.0253 | 0.9943 | - | | | | | | | | | |
| 4-O Mainland 1 | 0.9939 | 0.0395 | 0.0018 | - | | | | | | | | |
| 5-O Mainland 2 | 0.9725 | 0.9277 | 0.2317 | 0.4147 | - | | | | | | | |
| 6-O SR Grimness | 0.0054 | 0.9218 | 1.0000 | 0.0004 | 0.0653 | - | | | | | | |
| 7-O SR Windwick | 0.0766 | 0.9999 | 1.0000 | 0.0065 | 0.4943 | 0.9997 | - | | | | | |
| 8-O W Swartmill | 1.0000 | 0.1647 | 0.0088 | 0.9995 | 0.8676 | 0.0016 | 0.0305 | - | | | | |
| 9-O W Ness | 0.9984 | 0.9261 | 0.3066 | 0.7189 | 1.0000 | 0.1138 | 0.5545 | 0.9828 | - | | | |
| 10-Pyrenean Navarra | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | - | | |
| 11-Pyrenean Belle Île | 0.1915 | 0.0000 | 0.0000 | 0.9667 | 0.0006 | 0.0000 | 0.0000 | 0.3556 | 0.0124 | 0.0002 | - | |
| 12-Pyrenean France | 0.0009 | 0.8185 | 1.0000 | 0.0001 | 0.0128 | 1.0000 | 0.9984 | 0.0002 | 0.0412 | 0.0000 | 0.0000 | - |

O = Orkney, SR = South Ronaldsay, W = Westray.

**Appendix 10.** Analysis of Variance (ANOVA) of *Sorex minutus* skulls.

Parametric ANOVA post-hoc test (Tukey–Kramer) on CS for *S. minutus* skulls (pairwise *p* values).

|  | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|
| 1-Ireland Cloghan | - | | | | | | | | | |
| 2-Ireland Derry | 1.0000 | - | | | | | | | | |
| 3-Ireland Gweedore | 0.2629 | 0.4309 | - | | | | | | | |
| 4-O Mainland 1 | 0.7358 | 0.8753 | 1.0000 | - | | | | | | |
| 5-O Mainland 2 | 0.1609 | 0.3100 | 1.0000 | 1.0000 | - | | | | | |
| 6-O SR Grimness | 0.5370 | 0.7312 | 1.0000 | 1.0000 | 1.0000 | - | | | | |
| 7-O SR Windwick | 0.3144 | 0.4932 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | - | | | |
| 8-O W Ness | 0.0076 | 0.0188 | 0.9697 | 0.8653 | 0.8655 | 0.8242 | 0.9624 | - | | |
| 9-O W Swartmill | 0.1365 | 0.2513 | 1.0000 | 0.9993 | 1.0000 | 0.9993 | 1.0000 | 0.9959 | - | |
| 10-Pyrenean Navarra | 0.7195 | 0.8866 | 0.9956 | 1.0000 | 0.9958 | 1.0000 | 0.9977 | 0.4177 | 0.9642 | - |

O = Orkney, SR = South Ronaldsay, W = Westray.

Non-parametric ANOVA post-hoc test (Tamhane) on CS for *S. minutus* skulls (pairwise *p* values).

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|
| 1-Ireland Cloghan | - | | | | | | | | | |
| 2-Ireland Derry | 1.0000 | - | | | | | | | | |
| 3-Ireland Gweedore | 0.1864 | 0.2943 | - | | | | | | | |
| 4-O Mainland 1 | 0.9929 | 0.9995 | 1.0000 | - | | | | | | |
| 5-O Mainland 2 | 0.6075 | 0.7936 | 1.0000 | 1.0000 | - | | | | | |
| 6-O SR Grimness | 0.4125 | 0.6003 | 1.0000 | 1.0000 | 1.0000 | - | | | | |
| 7-O SR Windwick | 0.0573 | 0.0799 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | - | | | |
| 8-O W Ness | 0.0001 | 0.0001 | 0.9913 | 0.9983 | 0.9989 | 0.5594 | 0.7247 | - | | |
| 9-O W Swartmill | 0.0698 | 0.1114 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | - | |
| 10-Pyrenean Navarra | 0.8372 | 0.9602 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 0.1230 | 0.9997 | - |

O = Orkney, SR = South Ronaldsay, W = Westray.

**Appendix 11.** Multivariate Analysis of Variance (MANOVA) of *Sorex minutus* mandibles.

Parametric MANOVA post-hoc tests (Hotelling $T^2$) on shape variables among groups of *S. minutus* mandibles (pairwise *p* values).

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1-Ireland Cloghan | - | 0.2099 | 0.1667 | Fail | 0.0000 | 0.0035 | 0.0178 | 0.0000 | 0.0000 | 0.0080 | 0.0095 | 0.0000 |
| 2-Ireland Derry | 1.0000 | - | 0.1352 | 0.1351 | 0.0000 | 0.0001 | 0.0000 | 0.0000 | 0.0000 | 0.0006 | 0.0004 | 0.0000 |
| 3-Ireland Gweedore | 1.0000 | 1.0000 | - | Fail | 0.0000 | 0.0244 | 0.0017 | 0.0008 | 0.0001 | 0.0016 | 0.0027 | 0.0000 |
| 4-O Mainland 1 | Fail | 1.0000 | Fail | - | 0.0032 | Fail | Fail | Fail | Fail | 0.0347 | Fail | 0.0004 |
| 5-O Mainland 2 | 0.0000 | 0.0000 | 0.0003 | 0.2085 | - | 0.0017 | 0.0015 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 |
| 6-O SR Grimness | 0.2305 | 0.0041 | 1.0000 | Fail | 0.1102 | - | 0.4906 | 0.0029 | 0.0002 | 0.0285 | 0.0013 | 0.0001 |
| 7-O SR Windwick | 1.0000 | 0.0005 | 0.1125 | Fail | 0.0974 | 1.0000 | - | 0.0001 | 0.0003 | 0.0000 | 0.0014 | 0.0000 |
| 8-O W Swartmill | 0.0016 | 0.0000 | 0.0500 | Fail | 0.0001 | 0.1926 | 0.0065 | - | 0.0146 | 0.0000 | 0.0003 | 0.0000 |
| 9-O W Ness | 0.0019 | 0.0003 | 0.0037 | Fail | 0.0000 | 0.0164 | 0.0209 | 0.9633 | - | 0.0000 | 0.0005 | 0.0000 |
| 10-Pyrenean Navarra | 0.5261 | 0.0381 | 0.1074 | 1.0000 | 0.0000 | 1.0000 | 0.0007 | 0.0000 | 0.0000 | - | 0.0012 | 0.0000 |
| 11-Pyrenean Belle Île | 0.6297 | 0.0271 | 0.1782 | Fail | 0.0000 | 0.0884 | 0.0948 | 0.0183 | 0.0335 | 0.0769 | - | 0.0021 |
| 12-Pyrenean France | 0.0000 | 0.0000 | 0.0000 | 0.0243 | 0.0000 | 0.0063 | 0.0000 | 0.0000 | 0.0000 | 0.0005 | 0.1395 | - |

Above diagonal, Bonferroni uncorrected *p* values. Below diagonal, Bonferroni corrected *p* values. O = Orkney, SR = South Ronaldsay, W = Westray.

Non-parametric MANOVA post-hoc tests on shape variables among groups of *S. minutus* mandibles (pairwise *p* values).

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1-Ireland Cloghan | - | 0.0796 | 0.0452 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0023 | 0.0000 |
| 2-Ireland Derry | 1.0000 | - | 0.0871 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0006 | 0.0000 |
| 3-Ireland Gweedore | 1.0000 | 1.0000 | - | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0047 | 0.0000 |
| 4-O Mainland 1 | 0.0000 | 0.0000 | 0.0000 | - | 0.0000 | 0.0001 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0110 | 0.0000 |
| 5-O Mainland 2 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | - | 0.0044 | 0.0264 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 |
| 6-O SR Grimness | 0.0000 | 0.0000 | 0.0000 | 0.0066 | 0.2904 | - | 0.0123 | 0.0000 | 0.0000 | 0.0000 | 0.0002 | 0.0000 |
| 7-O SR Windwick | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 1.0000 | 0.8118 | - | 0.0000 | 0.0000 | 0.0000 | 0.0001 | 0.0000 |
| 8-O W Swartmill | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | - | 0.0000 | 0.0000 | 0.0000 | 0.0000 |
| 9-O W Ness | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | - | 0.0000 | 0.0000 | 0.0000 |
| 10-Pyrenean Navarra | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | - | 0.0001 | 0.0000 |
| 11-Pyrenean Belle Île | 0.1518 | 0.0396 | 0.3102 | 0.7260 | 0.0000 | 0.0132 | 0.0066 | 0.0000 | 0.0000 | 0.0066 | - | 0.0210 |
| 12-Pyrenean France | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 1.0000 | - |

Above diagonal, Bonferroni uncorrected *p* values. Below diagonal, Bonferroni corrected *p* values. O = Orkney, SR = South Ronaldsay, W = Westray.

**Appendix 12.** Multivariate Analysis of Variance (MANOVA) of *Sorex minutus* skulls.

Parametric MANOVA post-hoc tests (Hotelling T$^2$) on shape variables among groups of *S. minutus* skulls (pairwise *p* values).

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|
| 1-Ireland Cloghan | - | 0.1673 | 0.2792 | Fail | 0.0002 | 0.1180 | 0.1734 | 0.0028 | 0.0428 | 0.0003 |
| 2-Ireland Derry | 1.0000 | - | 0.6732 | Fail | 0.0023 | 0.0659 | 0.2971 | 0.0073 | 0.0041 | 0.0000 |
| 3-Ireland Gweedore | 1.0000 | 1.0000 | - | Fail | 0.0001 | 0.2050 | 0.2727 | 0.0012 | 0.0525 | 0.0002 |
| 4-O Mainland 1 | Fail | Fail | Fail | - | 0.3488 | Fail | Fail | Fail | Fail | 0.0379 |
| 5-O Mainland 2 | 0.0102 | 0.1026 | 0.0035 | 1.0000 | - | 0.0001 | 0.0044 | 0.0000 | 0.0000 | 0.0000 |
| 6-O SR Grimness | 1.0000 | 1.0000 | 1.0000 | Fail | 0.0029 | - | 0.7743 | 0.0485 | 0.1161 | 0.0000 |
| 7-O SR Windwick | 1.0000 | 1.0000 | 1.0000 | Fail | 0.2000 | 1.0000 | - | 0.1981 | 0.2553 | 0.0001 |
| 8-O W Ness | 0.1253 | 0.3292 | 0.0549 | Fail | 0.0000 | 1.0000 | 1.0000 | - | 0.6917 | 0.0000 |
| 9-O W Swartmill | 1.0000 | 0.1856 | 1.0000 | Fail | 0.0000 | 1.0000 | 1.0000 | 1.0000 | - | 0.0001 |
| 10-Pyrenean Navarra | 0.0130 | 0.0001 | 0.0093 | 1.0000 | 0.0000 | 0.0005 | 0.0064 | 0.0002 | 0.0043 | - |

Above diagonal, Bonferroni uncorrected *p* values. Below diagonal, Bonferroni corrected *p* values. O = Orkney, SR = South Ronaldsay, W = Westray.

Non-parametric MANOVA post-hoc tests on shape variables among groups of *S. minutus* skulls (pairwise *p* values).

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|
| 1-Ireland Cloghan | - | 0.0076 | 0.0190 | 0.0000 | 0.0000 | 0.0002 | 0.0000 | 0.0000 | 0.0000 | 0.0000 |
| 2-Ireland Derry | 0.3420 | - | 0.3101 | 0.0020 | 0.0004 | 0.0036 | 0.0001 | 0.0000 | 0.0000 | 0.0000 |
| 3-Ireland Gweedore | 0.8550 | 1.0000 | - | 0.0000 | 0.0000 | 0.0104 | 0.0002 | 0.0000 | 0.0000 | 0.0000 |
| 4-O Mainland 1 | 0.0000 | 0.0900 | 0.0000 | - | 0.0270 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 |
| 5-O Mainland 2 | 0.0000 | 0.0180 | 0.0000 | 1.0000 | - | 0.0002 | 0.0000 | 0.0000 | 0.0000 | 0.0000 |
| 6-O SR Grimness | 0.0090 | 0.1620 | 0.4680 | 0.0000 | 0.0090 | - | 0.1864 | 0.0000 | 0.0000 | 0.0000 |
| 7-O SR Windwick | 0.0000 | 0.0045 | 0.0090 | 0.0000 | 0.0000 | 1.0000 | - | 0.0000 | 0.0000 | 0.0000 |
| 8-O W Ness | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | - | 0.6730 | 0.0000 |
| 9-O W Swartmill | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 1.0000 | - | 0.0000 |
| 10-Pyrenean Navarra | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | - |

Above diagonal, Bonferroni uncorrected *p* values. Below diagonal, Bonferroni corrected *p* values. O = Orkney, SR = South Ronaldsay, W = Westray.