The LINNEAN SOCIETY of London BIOLOGICAL Journal of the Linnean Society

*Biological Journal of the Linnean Society,* 2016, ••, ••–••. With 4 figures.

# Post-glacial colonization of Europe by the wood mouse, *Apodemus sylvaticus*: evidence of a northern refugium and dispersal with humans

JEREMY S. HERMAN[1]*, FRÍðA JÓHANNESDÓTTIR[2], ELEANOR P. JONES[3], ALLAN D. MCDEVITT[4,5], JOHAN R. MICHAUX[6], THOMAS A. WHITE[2,7], JAN M. WÓJCIK[5] and JEREMY B. SEARLE[2]

[1]*National Museums of Scotland, Chambers Street, Edinburgh, EH1 1JF, UK*
[2]*Department of Ecology and Evolutionary Biology, Cornell University, Corson Hall, Ithaca, NY, 14853-2701, USA*
[3]*Fera Science, Sand Hutton, York, YO41 1LZ, UK*
[4]*Ecosystems and Environment Research Centre, School of Environment and Life Sciences, University of Salford, Salford, M5 4WT, UK*
[5]*Mammal Research Institute, Polish Academy of Sciences, 17-230, Białowieża, Poland*
[6]*Unité de génétique de la conservation, Institut de Botanique, Université de Liège, 4000, Liège, Belgique*
[7]*Lancaster Environment Centre, Lancaster University, Lancaster, LA1 4YQ, UK*

*Received 2 May 2016; revised 10 July 2016; accepted for publication 11 July 2016*

The wood mouse *Apodemus sylvaticus* is an opportunistic rodent that is found throughout most of the European mainland. It is present on many islands around the margins of the continent and in northern Africa. The species has been the subject of previous phylogeographical studies, although these have focussed on the more southerly part of its range. A substantial number of new samples, many of them from the periphery of the species' range, contribute to an exceptional dataset comprising 981 mitochondrial cytochrome *b* sequences. These new data provide sufficient resolution to transform our understanding of the survival of the species through the last glaciation and its subsequent re-colonization of the continent. The deepest genetic split that we found is in agreement with previous studies and runs from the Alps to central Ukraine, although we further distinguish two separate lineages in wood mice to the north and west of this line. It is likely that this part of Europe was colonized from two refugia, putatively located in the Iberian peninsula and the Dordogne or Carpathian region. The wood mouse therefore joins the growing number of species with extant populations that appear to have survived the Last Glacial Maximum in northern refugia, rather than solely in traditionally recognized refugial locations in the southern European peninsulas. Furthermore, the existence of a northern refugium for the species was predicted in a study of mitochondrial variation in a specific parasite of the wood mouse, demonstrating the potential value of data from parasites to phylogeographical studies. Lastly, the presence of related haplotypes in widely disparate locations, often on islands or separated by substantial bodies of water, demonstrates the propensity of the wood mouse for accidental human-mediated transport. © 2016 The Linnean Society of London, *Biological Journal of the Linnean Society*, 2016, **00**, 000–000.

KEYWORDS: cytochrome *b* – demography – Human introduction – mitochondrial DNA – molecular clock – phylogeography.

## INTRODUCTION

The wood (or field) mouse *Apodemus sylvaticus* (Linnaeus, 1758) is one of several European representatives from a genus that is distributed across the Palaearctic region (Musser & Carleton, 2005). The wood mouse shares much of its range with the closely-related yellow-necked mouse *Apodemus flavicollis* (Melchior, 1834), although it is found in many locations around the western margins of the

*Corresponding author. E-mail: J.Herman@nms.ac.uk

continent and in North Africa, whereas the yellow-necked mouse reaches further east into western Asia (Montgomery, 1999a, b; Amori *et al.*, 2008; Schlitter *et al.*, 2008). Although the wood mouse is often associated with temperate woodland, it is highly adaptable and opportunistic, and is found in a wide range of habitats, including forests, hedgerows, fields, grassland, gardens, and dunes (Flowerdew & Tattersall, 2008).

The phylogeography of this very numerous and widely distributed European species has been investigated previously in a number of studies (Michaux *et al.*, 2003; Michaux, Libois & Filippucci, 2005; Hoofer *et al.*, 2007; Lalis *et al.*, 2016), based on various series of mitochondrial cytochrome *b* (cyt *b*) gene sequences. Unsurprisingly, given the emphasis at that time on Mediterranean peninsulas as glacial refugia and as the source of populations that colonized northern Europe in the Holocene (Taberlet *et al.*, 1998; Hewitt, 1999), the earlier work was focused on the southern part of the continent and few data were related to the northern and more peripheral parts of the species' range (Michaux *et al.*, 2003). Subsequently, more northerly glacial refugia have been identified as the source of contemporary populations in several temperate European mammal species (Deffontaine *et al.*, 2005; Kotlík *et al.*, 2006; Vega *et al.*, 2010; Wójcik *et al.*, 2010; McDevitt *et al.*, 2012), and it is now widely accepted that the Mediterranean peninsulas were often harbours of endemic diversity, rather than sources for subsequent population expansion (Bilton *et al.*, 1998). In the present study, we have 454 new cyt *b* sequences from these more northerly and peripheral locations, whereas another recent study provides a large volume of data from North Africa (Lalis *et al.*, 2016), such that sampling now covers most of the species' range. This impressive dataset now includes 981 wood mouse cyt *b* sequences and 543 haplotypes.

In addition, the field of phylogeography has developed considerably since these earlier studies, and we have applied modern techniques and assumptions to the analysis of the data. In particular, there is now a radically altered understanding of the timing of phylogeographical events, following the realization that molecular clock rates are time dependent (Ho *et al.*, 2005), when measured over the recent timescales that apply to studies of intraspecific genetic variation. This has led to an appreciation of the need for calibrations that are directly based on the data themselves, or at least on equivalent data from other species, rather than deep interspecific splits that are associated with fossils of known age (Ho *et al.*, 2005, 2011; Ho & Larson, 2006). In the present study, we have used a mitochondrial cyt *b* clock rate that was directly estimated from 23 radiocarbon dated sub-fossil samples of another rodent, the common vole *Microtus arvalis* (Martínková *et al.*, 2013). These remains were obtained from archaeological sites in Orkney and were therefore deposited over a timescale similar to that likely to operate here and under the specific climatic background of the later Pleistocene and Holocene in Europe.

The wood mouse is present on numerous islands around the Atlantic coasts of Europe, including those of the British Isles, as well as those of the Mediterranean Sea (Corbet, 1978; Montgomery, 1999b). By contrast, the closely-related yellow-necked mouse is found in much of the wood mouse mainland European range, although it is recorded from far fewer island locations (Montgomery, 1999a). The prevalence of the wood mouse on islands, and its close association with human activities, led to the realization that it had been introduced, presumably by accident, to the islands around Britain during the Holocene (Corbet, 1961; Berry, 1969). Accordingly, the frequencies of non-metrical skeletal characters in island and mainland populations were used as a proxy for genetic variation, aiming to infer the source of the introductions of wood mice and other island mammals (Berry, Evans & Sennitt, 1967; Berry, 1969, 1973). The results of these studies appeared to implicate Viking settlers from Scandinavia in the introduction of these species to the islands around Britain, a conclusion that has been confirmed more recently with molecular data from the western house mouse *Mus musculus domesticus* Schwarz and Schwarz, 1943 (Searle *et al.*, 2009a; Jones *et al.*, 2012, 2013). These analyses of phenotypic characters share a common goal with many of the phylogeographical analyses that have subsequently been developed with the aim of placing genetic variation in its historical and geographical context.

Given the number of these island wood mouse populations, accidental human introductions may have played an important role in the dispersal of the species at the margins and, by implication, elsewhere in its range. Despite the earlier studies using non-metrical characters, this aspect of post-glacial colonization has received little attention in previous phylogeographical studies of the wood mouse, other than a recent study focusing on the North African range of the species (Lalis *et al.*, 2016). This may be a result of their bias towards the Mediterranean peninsulas and the more central European mainland (Michaux *et al.*, 2003, 2005; Hoofer *et al.*, 2007). In addition to range-wide European phylogeographical reconstruction, the present study therefore seeks to examine the role of human introductions in the post-glacial spread of the wood mouse.

## MATERIAL AND METHODS

Genomic DNA was extracted from 407 frozen or ethanol preserved tissue samples and from 47 preserved skins or dried soft tissue adhering to skulls, using commercially available extraction kits (DNeasy®; Qiagen). In the case of frozen or ethanol preserved tissues, the entire 1140-bp cyt *b* gene was generally amplified in a single polymerase chain reaction (PCR) using standard protocols and primers that were originally adapted for *Microtus* voles (Jaarola & Searle, 2002) but are also effective for other mammals including *Apodemus*. In a limited number of cases, a shorter segment of the cyt *b* gene (usually trimmed to 818 bp) was amplified using primers developed for *Apodemus* (Michaux *et al.*, 2003). For museum skins and other dried tissue, the cyt *b* gene was amplified using a protocol designed for this type of material, along with a suite of new or previously published primers (see Supporting information, Table S1) (Conroy & Cook, 1999; Michaux *et al.*, 2001; Jaarola & Searle, 2002; Wójcik *et al.*, 2010). Museum skins yield reduced amounts of DNA, which is also more fragmented; therefore, the gene was amplified in five overlapping fragments of approximately 230–360 bp. Negative controls, with no tissue in the DNA extraction and no template DNA in the PCR, were included in each procedure.

Sequencing was carried out in commercial facilities, using primers listed in the Supporting information (Table S1). For DNA amplified from frozen or ethanol preserved tissue samples, the two PCR primers and two newly designed internal primers were initially used, although only the two internal primers were used latterly. For DNA amplified from skins or other dried tissue, the PCR primers were used. Sequences for all new haplotypes reported in the present study have been deposited in the GenBank database (KX159497–KX159717) and, where available, voucher specimens are preserved in either the National Museums of Scotland or the Mammal Research Institute, Polish Academy of Sciences.

An alignment was prepared from the 454 new sequences that were generated in the course of the present study, together with 527 previously published sequences from the species, using SEAVIEW, version 4.5.4 (Gouy, Guindon & Gascuel, 2010). Details of all sequences, including source, sample material, geographical locality, GenBank and voucher reference numbers, are provided in the Supporting information (Table S2). Most of the previously published sequences are considerably shorter than the complete 1140-bp cyt *b* gene amplified in the present study; therefore, the lengths of sequences are also included. There has been some confusion in earlier literature (Michaux *et al.*, 2003, 2005; Lalis *et al.*, 2016) regarding data for some of the published sequences. The anomalies have been corrected in the course of the present study and the list of data in the Supporting information (Table S2) may be taken as being definitive for all the *A. sylvaticus* cyt *b* sequences that are available in the GenBank database at the time of writing.

Haplotypes were determined using ARLEQUIN, version 3.5 (Excoffier & Lischer, 2010) and subsequent visual checking of alignments with SEAVIEW, version 4.5.4. Sequences differing only by missing calls were initially defined as separate haplotypes, which were used in the phylogenetic reconstruction. These haplotypes are shown in the complete list of specimens (see Supporting information, Table S2). For geographical analysis (see below), sequences differing only by missing nucleotide data were defined as the same haplotype.

JMODELTEST, version 2.1.7 (Darriba *et al.*, 2012) was used to compare a range of GTR-nested nucleotide substitution models, with or without unequal base frequencies and a gamma distribution of rates across sites. The best model was selected and used in subsequent phylogenetic and population genetic analyses, based on decision theory (DT) or Bayesian information criterion (BIC), depending on the particular method employed in these analyses.

A maximum likelihood (ML) tree was inferred from the 543 haplotypes with PHYML, version 3.0 (Guindon *et al.*, 2010). The substitution model was Tamura–Nei+G (Tamura & Nei, 1993), a single transition and two transversion rates with gamma distribution across sites and unequal base frequencies, selected according to the DT (above). The outgroup used was *A. flavicollis*, which is included within the same subgenus *Sylvaemus* as *A. sylvaticus* (Michaux *et al.*, 2002; Suzuki *et al.*, 2008). Branch support was quantified by the Shimodaira–Hasegawa-like test (SH-aLRT) implemented in PHYML, version 3.0 (Guindon *et al.*, 2010), a nonparametric version of the approximate likelihood ratio test (aLRT) (Anisimova & Gascuel, 2006). The SH-aLRT is based on the Shimodaira–Hasegawa multiple tree comparison procedure (Shimodaira & Hasegawa, 1999). It makes use of intermediate likelihood values from the original ML tree search and, consequently, offers a fast alternative to bootstrap support values, which require the inference of many replicate ML trees (Anisimova *et al.*, 2011). The SH-aLRT has slightly more power and is almost as conservative as the standard nonparametric bootstrap procedure, at least for support thresholds down to 0.8 (Anisimova *et al.*, 2011); therefore, the interpretation of support values from the two methods will be similar.

Nuclear mitochondrial translocations (numts) were unwittingly included in previous phylogenetic

analyses of *Apodemus* cyt *b* gene sequences (Martin *et al.*, 2000; Reutter *et al.*, 2003; Suzuki *et al.*, 2008); however, their presence and characteristics have subsequently been established (Dubey *et al.*, 2009). The newly prepared alignment was therefore checked for anomalous sequences, containing nonsense codons, stop codons or shifts of the reading frame because these might represent numts. In addition, another ML tree was inferred, using the whole *A. sylvaticus* cyt *b* alignment together with outgroup sequences from four other members of the subgenus *Sylvaemus* and four *A. sylvaticus* numt sequences available from GenBank, aiming to ensure that our cyt *b* alignment did not contain any sequences similar to these previously identified numts.

Reticulating networks may be more effective than bifurcating trees with respect to recovering intraspecific patterns of genetic variation (Posada & Crandall, 2001). Minimum-spanning and median-joining algorithms (Bandelt, Forster & Röhl, 1999), both implemented in POPART, version 1.7 (http://popart.otago.ac.nz), were used to infer networks from all of the 981 cyt *b* sequences (see Supporting information, Table S2). The reticulating network was then compared with the ML tree to confirm the structure and to identify any additional information about the relationships among lineages.

The genetic structure, as identified in the ML tree, was tested by analysis of molecular variance (AMOVA), based on genetic distances within and among the mitochondrial lineages, using ARLEQUIN, version 3.5. Percentages of variation and $\Phi_{ST}$ were calculated and their significance tested with 10 000 permutations of individuals among lineage populations. In addition, $\Phi_{ST}$ was calculated for each pairwise comparison of lineage populations, with significance again tested using 10 000 permutations.

Genetic variation in lineage populations was examined using a range of analyses available in DNASP, version 5.10.01 (Librado & Rozas, 2009). Nucleotide diversity ($\pi$), the average number of pairwise differences (per site) between sequences, was calculated for each lineage and its SD was computed as the root of the variance. Two test statistics, *D* (Tajima, 1989) and $F_S$ (Fu, 1997) were also calculated for each lineage to identify episodes of possible recent demographic expansion in them. The significance of Tajima's *D* statistic was determined from published confidence limits with a $\beta$-distribution (Tajima, 1989), whereas the significance of Fu's $F_S$ was determined using null distributions obtained from 1000 coalescent simulations.

The frequency distributions of pairwise nucleotide site differences (mismatches) were obtained for each lineage using ARLEQUIN, version 3.5, aiming to determine whether the population had passed through a recent demographic expansion. A sudden demographic expansion leads to a unimodal mismatch distribution that can be described by a model with three parameters, which are composed of the mutation rate together with the time since expansion ($\tau$), and the population size before ($\theta_0$) and after ($\theta_1$) the event (Rogers & Harpending, 1992). These were calculated using a generalized least-square approach (Schneider & Excoffier, 1999) and the expected distribution was compared with the observed one graphically and by computing the sum of squared deviations (SSD) between the two distributions and the raggedness index (*r*). Significance of the SSD and *r*, together with confidence intervals for the $\tau$ parameter, were obtained by re-calculating them for each distribution obtained from 10 000 coalescent simulations that used the original estimated parameters of the demographic expansion. An indication of the relative time of expansion was obtained for each lineage, by comparing the values of nucleotide diversity ($\pi$) and the mismatch $\tau$ parameter from each of them.

More sophisticated demographic analyses were carried out using Bayesian Markov chain Monte Carlo (MCMC) simulations in BEAST, version 2.3.1 (Bouckaert *et al.*, 2014). Individual coalescent genealogies and skyline models (Drummond *et al.*, 2005) of the effective population size for female voles belonging to each of the six identified mitochondrial lineages were co-estimated using a shared model of sequence evolution. Separate genealogies and demographic models are appropriate for these distinct female lineage populations, which are presumed to be independent and the result of distinct population expansions, because this takes account of genetic structure within the overall population of the species. The number of groups in the skyline model for each lineage was based on the (unequal) number of sequences within them. The shared substitution model used was the Tamura–Nei+G (Tamura & Nei, 1993), as selected using the BIC in JMODELTEST, version 2.1.7. Base frequencies, kappa values for transition/transversion rates, and the $\alpha$ parameter of the gamma distribution of rates were all estimated along with the other parameters of the model. The molecular clock rate was fixed at $3.27 \times 10^{-7}$ substitutions site$^{-1}$ year$^{-1}$, which was previously estimated from radiocarbon dated sequences of another rodent, the common vole *Microtus arvalis* (Martínková *et al.*, 2013). A strict molecular clock was used, given that the sequences are from a single species and there should be little variation in branch rates. Prior parameter distributions are shown in the Supporting information (Table S3) and all simulations were repeated without sequence data to test the joint distributions of parameters obtained with

the priors alone and to ensure that the results were not unduly influenced by these.

Posterior parameter distributions were obtained from eight separate MCMC chains that were each run for 100 million generations, using different random seeds, with the first 10 million generations removed as burn-in. Log files from these eight simulations were examined using TRACER, version 1.6 (Rambaut *et al.*, 2014), to check for convergence, based on the posterior distributions and the traces of the values for each parameter. The effective sample size (ESS) for each parameter was at least 200, which is generally considered to be a sufficient sample of the posterior distribution. The log files were combined using LOGCOMBINER in the BEAST, version 2.3.1 package. The posterior distribution, value trace and ESS for each parameter was then examined in the combined file using TRACER, version 1.6 (Rambaut *et al.*, 2014). TRACER, version 1.6 was also used to derive the Bayesian Skyline Plot (BSP), effective female population size over time, for each lineage.

For haplotypes found in more than one geographical location, pairwise geodesic distances (World Geodetic System: WGS84) were calculated between each of these locations. Locations within a single land mass or island archipelago were collapsed to their mean (for two locations) or centroid coordinates. Connections between locations with identical haplotypes were mapped, where these involved sea crossings. Calculations were carried out using R, version 3.2.2 (Hijmans, 2015; R Core Team, 2015) and mapping with QGIS, version 2.0.1 (QGIS Development Team 2004–2014), using *OpenStreetMap* coastline data (http://www.openstreetmap.org).

## RESULTS

The alignment contained 981 cyt *b* sequences (see Supporting information, Table S2) comprising 543 haplotypes, as defined above. There were 521 haplotypes after collapsing those distinguished only by missing nucleotide data. Most of the species' mainland European range (Montgomery, 1999b) has now been sampled, with substantial coverage of peripheral regions to the west, north, and east, including Iberia, France, Britain, Scandinavia, and Poland. Many of the islands on the Atlantic are now included, with samples from Iceland, Ireland, and 45 of the islands around Britain. In addition, many sequences from North Africa have recently become available (Lalis *et al.*, 2016). Remaining shortcomings in the coverage are largely confined to the south-eastern part of the wood mouse range and the Mediterranean islands.

There was no evidence for any *numts* among the sequences investigated in the present study. None of them contained any nonsense codons, nucleotide insertions or deletions that would shift the reading frame, and there were no internal stop codons. All of the wood mouse sequences clustered together in a highly supported (SH-aLRT 1.00) clade within the ML tree that was inferred for the whole alignment together with four *Sylvaemus* outgroup and four wood mouse *numt* sequences from GenBank (see Supporting information, Fig. S1). Furthermore, the wood mouse clade was less closely associated with the *numt* sequences than with the outgroup sequences, both of which also clustered into monophyletic groups.

The 543 haplotypes clustered into six distinct lineages within the ML tree (Fig. 1) and these have been named here according to the geographical distributions of the sampled sequences (Fig. 2). The *south-eastern*, *Sicilian*, and *African* lineages were previously identified in other studies (Michaux *et al.*, 2003, 2005; Lalis *et al.*, 2016). The *central* and *peripheral* lineages are newly identified here, having previously been considered a single homogeneous group (subclade 2b in Michaux *et al.*, 2003). The *Channel Island* lineage comprises only two new haplotypes (based on five sequences) from the islands of Sark and Guernsey, off the northern coast of France. A well-supported (SH-aLRT 0.97) node connects the *Channel Island* and *African* lineages. Five of the six lineages had strong SH-aLRT support (0.86–0.97) in the ML tree (Fig. 1). The central lineage was less well-supported, with SH-aLRT of 0.78, which is just below the threshold where the SH-aLRT begins to diverge and become less conservative, albeit remaining substantially more powerful, than the standard bootstrap (Anisimova *et al.*, 2011).

The 543 haplotypes clustered into five distinct haplogroups in the minimum-spanning and median-joining networks, which were very similar to each other (Fig. 3). The *central* and *peripheral* haplogroups were very closely related to each other, as were the *south-eastern* and *Sicilian* haplogroups. Of the *central* and *peripheral* haplogroups, the former was more closely related to the *south-eastern* and *Sicilian* haplogroups. The *African* haplogroup was clearly associated with the *central* haplogroup and included the two haplotypes from the geographically distant Channel Islands of Sark and Guernsey.

An AMOVA, using genetic distances within and among the six lineage populations, showed that most of the variation was attributable to the divergence between the lineages ($\Phi_{ST} = 0.770$, bootstrap 0.05/99.95 percentile range 0.643–0.846). Pairwise fixation indices for combinations of the six lineages also lend support to the identified genetic structure (Table 1).

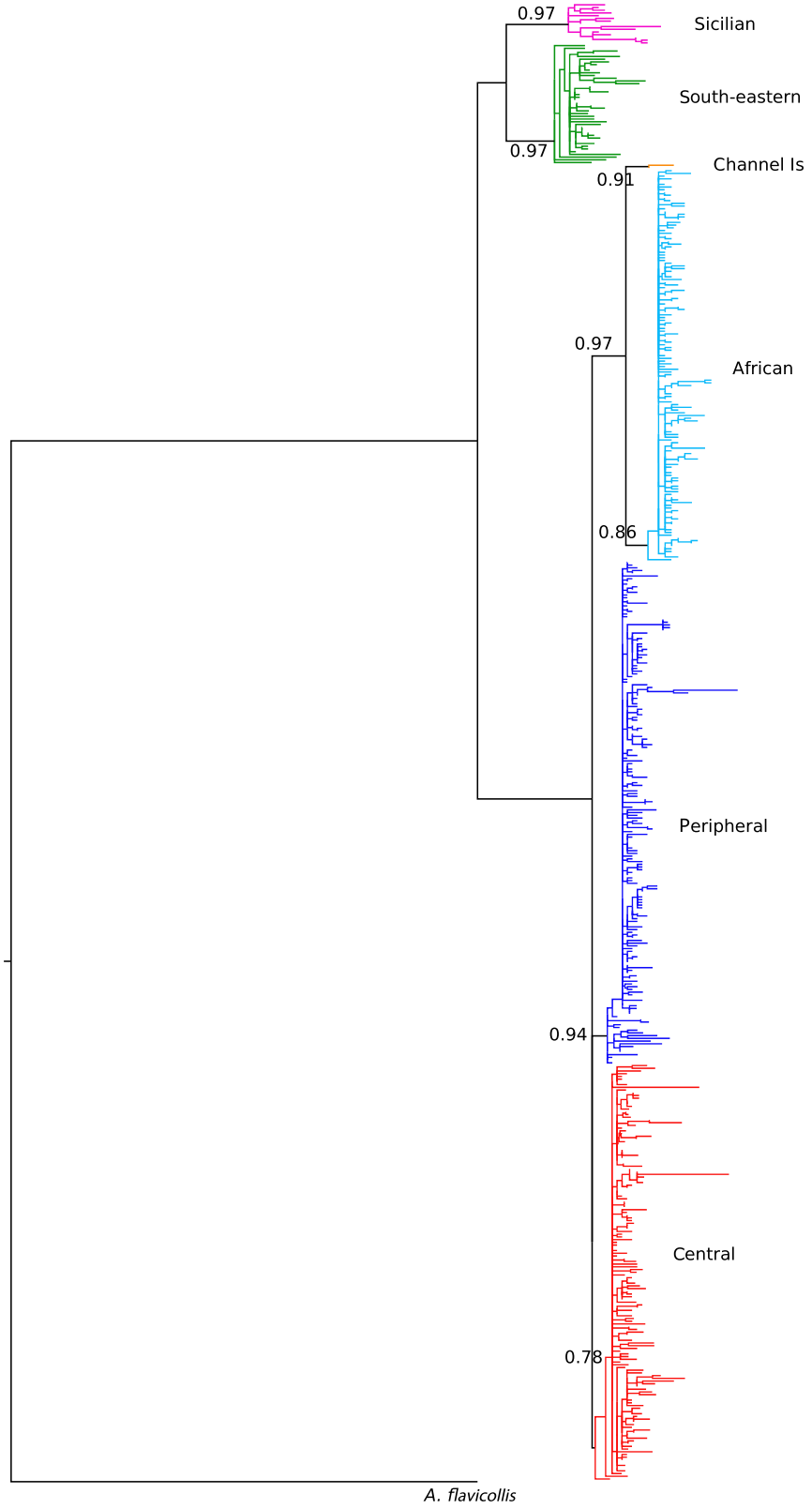


**Figure. 1.** Maximum likelihood tree for 543 wood mouse cytochrome *b* (cyt *b*) haplotypes. Tree inferred with PHYML, version 3.0, with support (SH-aLRT) only shown for identified lineages and deeper splits.

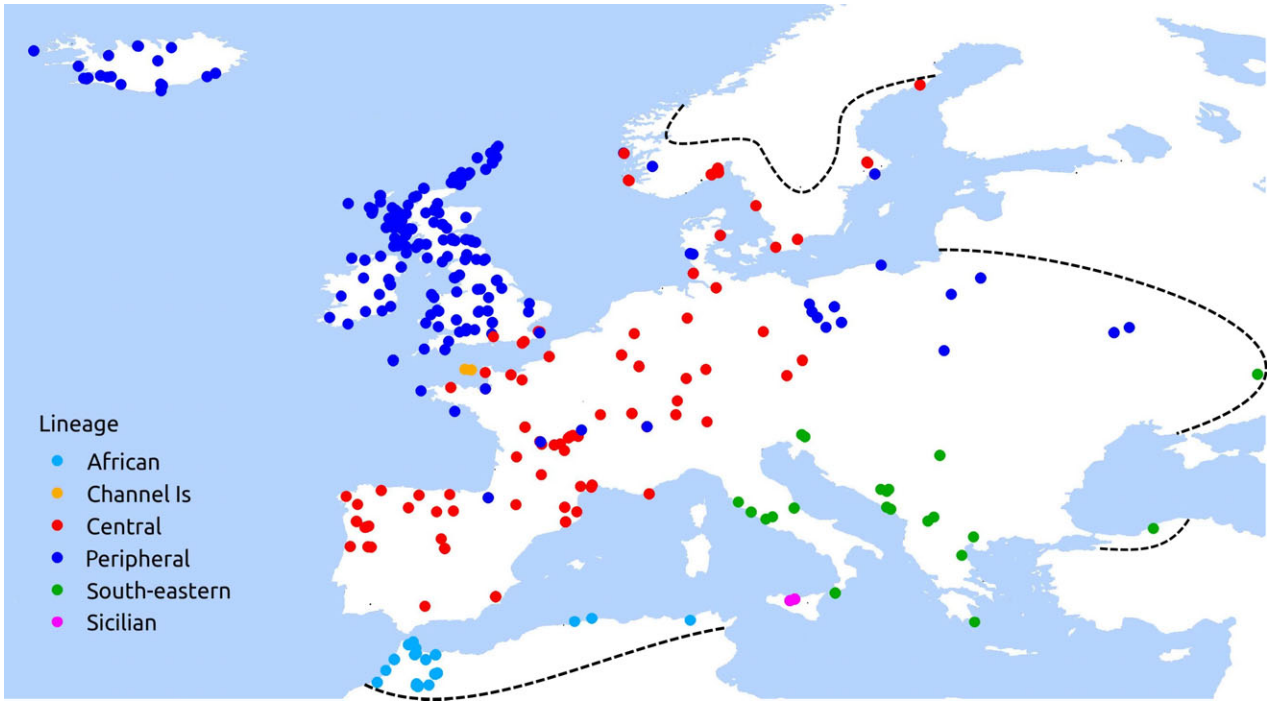


**Figure. 2.** Locations of samples from the six wood mouse cytochrome *b* lineages. Colours refer to lineages identified by maximum likelihood (ML) tree inference and other analyses. Coastline data from *OpenStreetMap* (http://www.openstreetmap.org). Limits of species range (dashed line) *sensu* Schlitter *et al.* (2008).

These ranged from 0.603 to 0.894 and all of them were significant based on 10 000 permutations of individuals among groups.

There is a clear genetic split in the data, which in mainland Europe is geographically aligned from the Alps to central Ukraine (Fig. 2). This genetic split is apparent in both the ML tree, where it is represented by a deep node within the genealogy (Fig. 1) and in the network, where the split is represented by a branch with multiple mutational steps (Fig. 3). The split is also apparent from the higher fixation indices for pairs that include either the *south-eastern* or *Sicilian* lineage, along with one of the other lineages. The continent to the south of this geographical divide, from the Italian and Balkan peninsulas to Ukraine, is occupied by the *south-eastern* lineage, together with the more closely-related and endemic *Sicilian* lineage, whereas, to the north and west, the remainder of Europe is occupied by the *central* and *peripheral* lineages, together with the *Channel Island* lineage (Fig. 2). The *central* lineage reaches from the Iberian peninsula in the west to the southern part of the Fennoscandian peninsula in the north. The *peripheral* lineage is found in predominantly more peripheral locations that extend from Ukraine and Poland in the east to the western European mainland, Scandinavia, the British Isles, and Iceland. There is also one more lineage, the *African*, found only in the North African part of the species' range.

Nucleotide diversity ($\pi$), an indicator of the relative age of the clade in question, is shown for each lineage in Table 2. The nucleotide diversity is higher in the *Sicilian* lineage (0.01295) than in the *south-eastern*, *central* or *peripheral* ones (0.00468–0.00597), whereas the *African* and the *Channel Island* lineages both have lower diversity (0.00366 and 0.00263).

The graphs of mismatch distributions for the *central*, *peripheral*, *African*, *south-eastern*, and *Sicilian* lineages all show the characteristic unimodal pattern that is indicative of a recent demographic expansion. However, the observed distributions of the latter two, and especially the *Sicilian* lineage, are not so close to the expected sudden expansion distributions with the model parameters (see Supporting information, Fig. S2). The mismatch distribution for the *Channel Island* lineage is bimodal, although the sample size for this group ($N = 5$) is very small. The SSD from the model distributions are low (0.001–0.014) for all of the lineages other than the *Channel Island* one (0.365), as expected after a recent demographic expansion, and none of the respective *P*-values indicate a significant departure from the model of recent expansion (Table 2). Similarly, the low values (0.005–0.028) of the raggedness indices (*ri*) for all but the *Channel Island* lineage (0.880) are

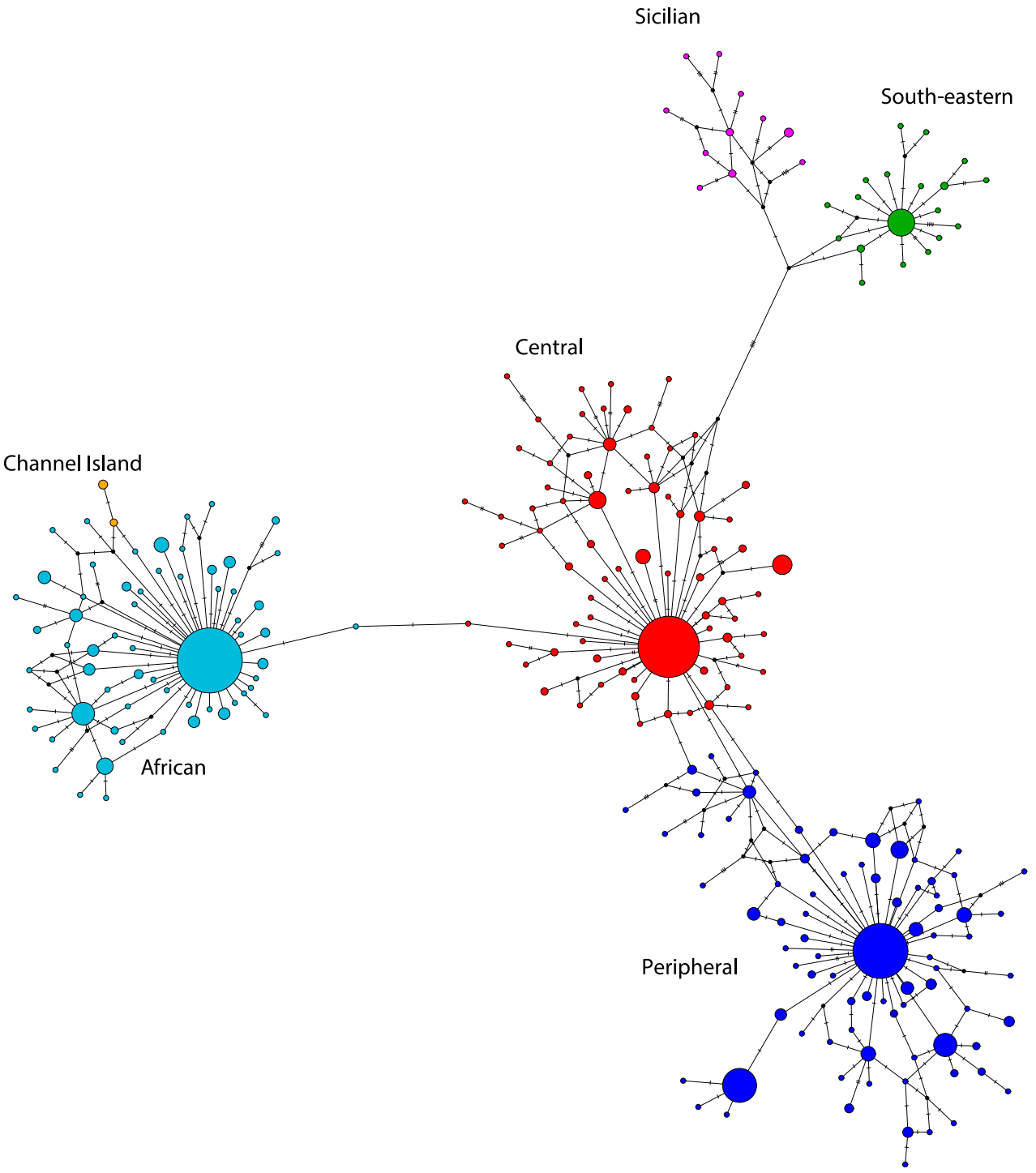


**Figure. 3.** Median-joining network for 981 wood mouse sequences, obtained with POPART, version 1.7. Area of circle represents frequency of haplotype and mutational steps are indicated by hatch marks across branches. Black dots indicate inferred intermediate haplotypes.

concomitant with recent expansions. The associated *P*-value for the *African* lineage shows a significant departure from the expansion model.

The mismatch Tau parameter ($\tau$), representing time since the onset of expansion, was similar for the *central*, *peripheral*, and *south-eastern* lineages

**Table 1.** Fixation indices ($\Phi_{ST}$) calculated from genetic distance in all pairwise combinations of lineages

| | African | Central | Channel Island | Peripheral | South-eastern |
|---|---|---|---|---|---|
| Central | 0.68884 | | | | |
| Channel Island | 0.71494 | 0.71941 | | | |
| Peripheral | 0.72329 | 0.60346 | 0.76854 | | |
| South-eastern | 0.89389 | 0.85019 | 0.84730 | 0.86032 | |
| Sicilian | 0.89367 | 0.84557 | 0.78516 | 0.86087 | 0.67214 |

All indices were significant ($P < 0.05$), based on 10 000 permutations of individuals among lineage populations.

**Table 2.** Genetic variation within the six wood mouse cytochrome *b* (cyt *b*) lineages

| | *n* | π | *D* | *P* | $F_S$ | *P* | Tau (τ) | 95% CI (τ) | SSD | *P* | *ri* | *P* |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| African | 298 | 0.00366 | −2.635 | < 0.001 | −232.616 | 0.000 | 2.875 | 2.191–5.168 | 0.001 | 0.090 | 0.028 | 0.016 |
| | | *0.00015* | | | | | | | | | | |
| Central | 274 | 0.00597 | −2.350 | < 0.01 | −148.553 | 0.000 | 5.395 | 2.977–7.139 | 0.001 | 0.549 | 0.010 | 0.567 |
| | | *0.00031* | | | | | | | | | | |
| Channel Island | 5 | 0.00263 | 1.686 | > 0.10 | 3.526 | 0.970 | 6.068 | 0.000–77.737 | 0.365 | 0.037 | 0.880 | 0.092 |
| | | *0.00077* | | | | | | | | | | |
| Peripheral | 342 | 0.00468 | −2.525 | < 0.001 | −207.570 | 0.000 | 4.000 | 3.275–6.621 | 0.002 | 0.102 | 0.013 | 0.343 |
| | | *0.00023* | | | | | | | | | | |
| South-eastern | 47 | 0.00506 | −2.330 | < 0.01 | −34.059 | 0.000 | 5.000 | 3.637–17.842 | 0.006 | 0.211 | 0.005 | 0.497 |
| | | *0.00046* | | | | | | | | | | |
| Sicilian | 15 | 0.01295 | −1.135 | > 0.10 | −6.577 | 0.004 | 8.000 | 5.562–20.121 | 0.014 | 0.160 | 0.015 | 0.755 |
| | | *0.00116* | | | | | | | | | | |

Nucleotide diversity (π), with SD in italics below; Tajima's *D*, with significance *P* from β-distribution; Fu's $F_S$, with significance *P* from 1000 coalescent simulations; mismatch Tau (τ) with 95% confidence interval, sum of squared deviations from model distribution (SSD) and Harpending's raggedness index (*ri*), with associated *P*-values derived from 10 000 coalescent simulations.

(4.000–5.395) but somewhat higher (8.000) in the *Sicilian* lineage and lower (2.875) in the *African* lineage (Table 2). This is broadly in accordance with the relative ages of the lineages from nucleotide diversity. The Tau parameter for the *Channel Island* lineage is high (6.068), despite its relatively low nucleotide diversity, although the sample size is very small ($N = 5$). In general, the widths of the 95% confidence intervals for the Tau parameter appear to reflect the sample size (Table 2). Tajima's *D* and Fu's $F_S$ statistics were negative for all lineages other than the *Channel Island* one (Table 2). These tests are more sensitive to demographic expansion than the mismatch *P*-values and all of the negative test statistics were significant, other than Tajima's *D* for the *Sicilian* lineage, which is of earlier origin according to the nucleotide diversity (π). There was no signal of recent expansion for the *Channel Island* lineage, although the sample size was very small.

With the previously estimated rodent cyt *b* clock rate ($3.27 \times 10^{-7}$ substitutions site$^{-1}$ year$^{-1}$; Martínková *et al.*, 2013), the times to the most recent common ancestor (tMRCAs) of the *central*, *peripheral*, and *south-eastern* lineages (estimated with Beast) are quite similar, with medians ranging from 16.4 to 22.3 kya (Table 3), as was the case for their Tau parameter values (Table 2). This time span is broadly coincident with the cold period around the Last Glacial Maximum (LGM; approximately 21.0 kya; Mix, Bard & Schneider, 2001) of the most recent (Weichselian) glaciation of Europe, just before the rapid increase in temperature that marked the beginning of the Bølling–Allerød interstadial (approximately 14.7 kya; Steffensen *et al.*, 2008). This is consistent with expansion of the three lineages out of three separate late glacial refugia. The tMRCAs for the *African* lineage (median 7.4 kya) and *Channel Island* lineage (median 5.6 Kya) are much more recent. The *African* lineage has a low value for the Tau parameter, in line with the more recent tMRCA, whereas the *Channel Island* lineage does not. Finally, the tMRCA for the *Sicilian* lineage is much earlier than the others (median 29.0 kya), as reflected in the higher value of its Tau parameter,

**Table 3.** Time to most recent common ancestor (tMRCA) for each mitochondrial lineage population

| Lineage | 95% HPD lower (kya) | Median (kya) | 95% HPD upper (kya) |
|---|---|---|---|
| African | 5.137 | 7.430 | 11.376 |
| Central | 9.404 | 22.254 | 37.355 |
| Channel Island | 1.339 | 5.591 | 11.543 |
| Peripheral | 8.689 | 16.363 | 32.252 |
| South-eastern | 13.779 | 19.868 | 29.079 |
| Sicilian | 19.628 | 29.018 | 40.483 |

Median and 95% highest posterior density (HPD) range of times, obtained from 540 million post-burn-in genealogy samples in BEAST, version 2.3.1. Trees calibrated with intraspecific rodent cytochrome *b* (cyt *b*) clock rate ($3.27 \times 10^{-7}$ substitutions site$^{-1}$ year$^{-1}$) estimated from radiocarbon dated samples of *Microtus arvalis* (Martínková *et al.*, 2013).

indicating that its most recent population expansion was initiated prior to the LGM.

The Bayesian skyline models, derived from the individual coalescent genealogies for each lineage, show the effective female population size ($N_{ef}$) from the median tMRCA of the respective lineage to the present (Fig. 4). Although the *south-eastern*, *central*, and *peripheral* lineages appear to have originated during the cold stage towards the end of the Weichselian glaciation, the demographic expansion of the *south-eastern* lineage was earlier than that of the other two lineages, approximately 14–15 kya, which coincides with the sudden and rapid warming of the climate approximately 14.7 kya (Steffensen *et al.*, 2008). However, the *central* and *peripheral* lineage populations did not expand until approximately 8–9 Kya, following the establishment of the Holocene interglacial, approximately 10.7 kya (Steffensen *et al.*, 2008). Demographic expansion of the *African* lineage closely followed its time of origin, beginning approximately 7 kya, as expected for an introduction to a new location. There is no signal of demographic expansion in the *Channel Island* lineage population, although the sample size was very small ($N = 5$). The population of the *Sicilian* lineage underwent a relatively gradual expansion, beginning sometime after its estimated time of origin, and coinciding with the period when the climate changed from its LGM minimum (approximately 21 kya) to its Holocene maximum (after 10 kya).

Although the overall genetic structure was clearly defined (Figs 1 and 3), and most of the variation was partitioned between the six lineages (Table 1), there was no discernible geographical structure within them. However, 68 different haplotypes were found in more than one geographical location (see Supporting information, Table S2) and pairwise geodesic distances between locations with the same haplotype ranged up to 2331 km. Of the 68 shared haplotypes, 14 were present on more than one land mass (see Supporting information, Table S4). Three of these haplotypes, all of them from the *central* lineage, were present on both the European mainland and the Fennoscandian peninsula, which were connected by land bridges between 12.7 and 13.1 kya and from 10.3 to 12.1 kya, prior to the formation of the Baltic Sea (Herman *et al.*, 2014; based on Björck, 1995a, b). Another two of the shared haplotypes were found on closely adjacent islands that might have been united with each other at some point after the last glaciation: Westray with mainland Orkney; Walney with Piel and Sheep Islands off north-western England. However, for the 12 remaining shared haplotypes, connections between their current locations would have required sea crossings, given the consistently high post-glacial sea levels around north-western Europe and the British Isles (Sturt, Garrow & Bradley, 2013). All of these potential overseas connections are between the British Isles, Iceland, and the southern part of the Fennoscandian peninsula (see Supporting information, Fig. S3).

## DISCUSSION

Our new phylogeographical data confirm the earlier finding (Michaux *et al.*, 2003) that the main genetic split in the wood mouse runs from the Alps to central Ukraine (Fig. 2). However, the higher resolution of the larger dataset, especially in the northern and marginal parts of the range, demonstrates that the main genetic group to the north and west of the split (subclade 2b of Michaux *et al.*, 2003) comprises two distinct and well-supported lineages, named here as the *central* and the *peripheral*, which have clearly different but overlapping distributions.

The identification of these two lineages gives rise to the most important new insight of the present study: that the mainland European population of the wood mouse has emerged from a combination of three glacial refugia, two of them somewhere in south-eastern and south-western Europe, and the third in a more northerly location. This interpretation is based on the present distributions of the *south-eastern* lineage, the *central* lineage that ranges from the Iberian peninsula north to Scandinavia, and the *peripheral* lineage that is confined to more northerly parts of the continent (Fig. 2). Unfortunately, the data are insufficient to give any indication of the precise locations of these putative glacial refugia. The existence of two southern refugia, in

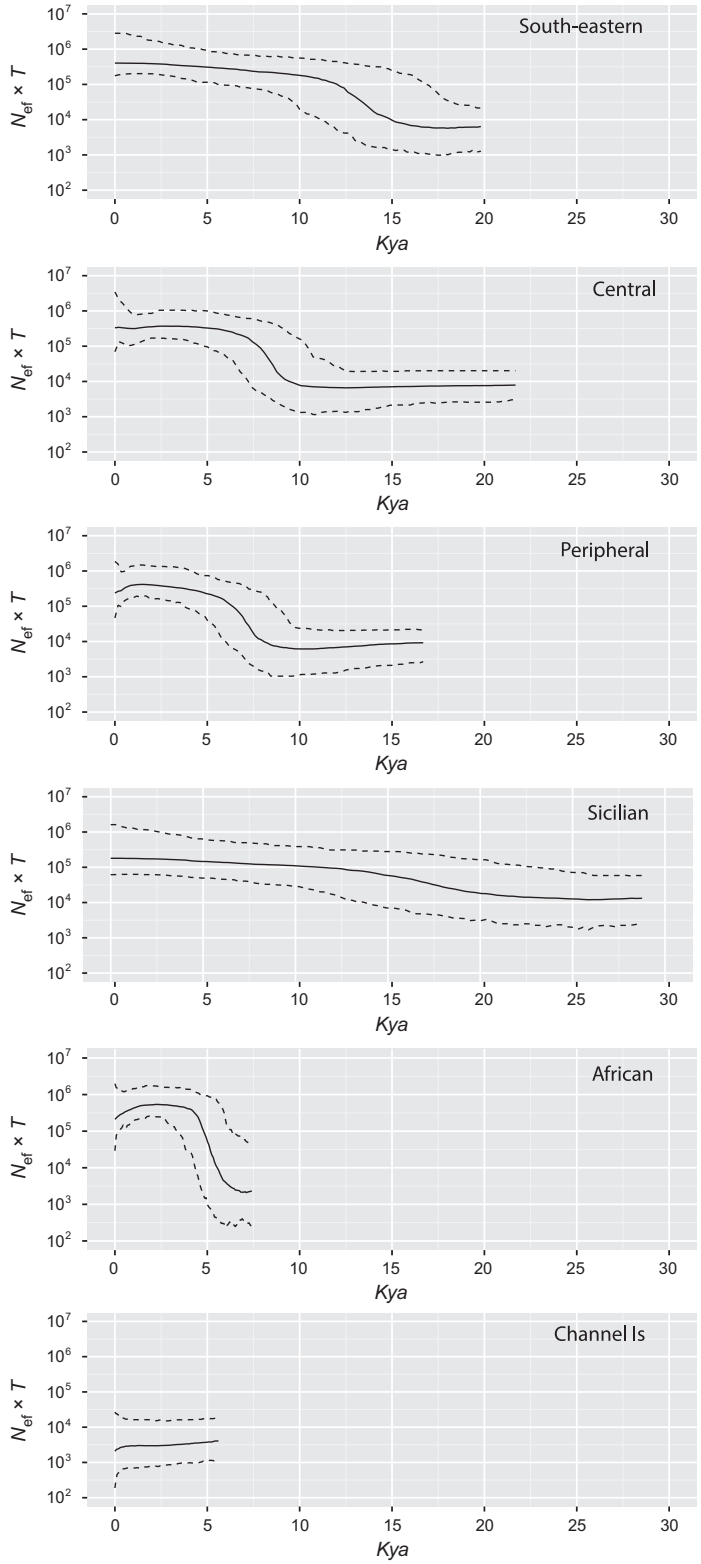


**Figure. 4.** Bayesian skyline plots showing effective female population size with time, for each lineage. Effective female population size ($N_{ef}$), multiplied by mean generation time ($T$), in years. Solid line is median and dashed lines are 95% highest posterior density (HPD) limits. $N_{ef} \times T$ plotted on log scale for clarity and truncated to median estimate of time to the most recent common ancestor.

Iberia and south-eastern Europe, was inferred in earlier studies of the wood mouse (Michaux *et al.*, 2003, 2005), although the presence of a northern refugium was not established. However, as stated earlier, the prevalence of northerly refugia was not generally recognized until relatively recently, and the phylogeographical pattern inferred here, with survival of lineages in a combination of southern and northern refugia, has now been found in a number of Eurasian rodents and other small mammals (Deffontaine *et al.*, 2005; Vega *et al.*, 2010; McDevitt *et al.*, 2012).

Although the presence of a northerly refugium has not previously been identified from genetic variation in the wood mouse itself, the existence of this refugium was predicted in a comparative phylogeographical study of the wood mouse and one of its parasites (Nieberding *et al.*, 2004, 2005). The nematode *Heligmosomoides polygyrus* (Dujardin, 1845) is a direct (without intermediate host) and specific endoparasite of the wood mouse, and it is therefore expected that genetic variation in their naturally occurring populations would be closely aligned. Refugia for both species were initially considered to be located in southern European peninsulas (Nieberding *et al.*, 2004). However, additional genetic structure was subsequently identified in populations of *H. polygyrus* cyt *b*, based on samples from Ireland and Denmark, and this was attributed to the presence of an unidentified refugium in a more northerly location (Nieberding *et al.*, 2005). This highlights the potential value of genetic data from parasites, particularly when data from the host are lacking or variation is limited by some factor such as introgression or selection.

The presence of northern and southern refugia may contribute to a general phylogeographical process that has been proposed, whereby northern or outlying locations are now occupied by the first lineages to colonize the continent, whereas more southern or central parts of the range are occupied by other lineages that have subsequently replaced them (Searle & Wilkinson, 1987; Piertney *et al.*, 2005; Searle *et al.*, 2009b). This process, and the resulting pattern of central and peripheral lineages, may be attributed to specific adaptive features of the lineages (McDevitt *et al.*, 2012; Kotlík *et al.*, 2014). Such a model might apply to the wood mouse, given the distributions of the *peripheral* and *central* lineages, which we have therefore named accordingly. If this were the case, the *peripheral* lineage may have been first to colonize central and northern Europe, but has subsequently been replaced by the *central* lineage in much of north-western and central Europe, although some 'relict' individuals remain scattered there. The median tMRCAs of the *peripheral* and *central* lineages (16.363 and 22.254 kya, respectively) (Table 3) appear to conflict with this interpretation, although these dates reflect the coalescence of each lineage whereas their demographic expansions were more closely aligned (beginning after 10 kya) (Fig. 4). If the *central* lineage began to expand contemporaneously from a more southerly location than the *peripheral* lineage, then it is conceivable that it would follow in the wake of latter and only partially replace it.

It has been demonstrated consistently, using directly and indirectly dated molecular data, that molecular clock rates are time=dependent, over the timescales that apply to studies of intraspecific genetic variation such as the present one (Ho *et al.*, 2005, 2011, 2015; Ho & Larson, 2006). In the absence of dated ancient DNA from the wood mouse itself, the need for calibration from contemporaneous events was met here by the application of a cyt *b* clock rate that was estimated using directly radiocarbon dated post-glacial remains from another rodent (Martínková *et al.*, 2013). Similar clock rates to the one used in the present study ($3.27 \times 10^{-7}$ substitutions site$^{-1}$ year$^{-1}$) have been inferred for another species of rodent in Europe, *Microtus agrestis* (Linnaeus, 1761), by comparing mitochondrial genetic variation with external events (Herman & Searle, 2011; Herman *et al.*, 2014), and such clock rates have also been applied to house mouse colonization (Rajabi-Maham, Orth & Bonhomme, 2008; Macholán *et al.*, 2012). Notably, similar clock rates have been obtained with dated ancient DNA from a wide range of mammals and birds (Ho, Kolokotronis & Allaby, 2007).

That rate is between two and three times higher than the rates ($0.9–1.2 \times 10^{-7}$ substitutions site$^{-1}$ year$^{-1}$) that were estimated for *A. sylvaticus* cyt *b* in the most recent phylogeographical study of the species (Lalis *et al.*, 2016). However, these were estimated from inferred splits between *A. sylvaticus*, *A. flavicollis*, and *Apodemus mystacinus*. Although those estimates were obtained from variation at third codon positions alone, this may not have been sufficient to account for all of the rate decay. The rate used here is also two times higher than recently estimated cyt *b* clock rates for *Apodemus argenteus* ($1.1–1.6 \times 10^{-7}$ substitutions site$^{-1}$ year$^{-1)}$ and *Apodemus speciosus* ($1.2–1.7 \times 10^{-7}$ substitutions site$^{-1}$ year$^{-1}$), obtained by relating the timing of demographic expansions in the Japanese archipelago to post-glacial climate (Suzuki *et al.*, 2015). This approach of calibration to external events has been applied elsewhere, such as with river catchments (Burridge *et al.*, 2008) and land bridges (Herman & Searle, 2011; Herman *et al.*, 2014), and therefore, as an *Apodemus* rate, may be more reliable than the

*Microtus* rate that we used. We have therefore considered the effect on our results, had we applied a clock rate ($1.4 \times 10^{-7}$ substitutions site$^{-1}$ year$^{-1}$) based on that obtained by external calibration of genetic variation in *Apodemus* (Suzuki *et al.*, 2015).

The directly dated *Microtus* molecular clock calibration leads to the following scenario:

The three mainland European lineages of wood mouse all originated around the time of the LGM (approximately 21.0 kya), presumably within bottlenecked populations that survived the LGM in three separate refugia. The subsequent demographic expansions of these three lineages are evident from the mismatch distributions (see Supporting information, Fig. S2), neutrality statistics (Table 2), and the skyline plots of population size (Fig. 4). Such expansions are, of course, expected in populations that must have occupied their present ranges in response to the climatic changes after the last glaciation. The demographic expansion of the *south-eastern* lineage began when the climate rapidly warmed at the beginning of the Bølling–Allerød interstadial (approximately 12.9 to 14.7 kya), which marked the end of the Weichselian glaciation. Presumably this lineage was soon able to occupy most of its present range (Fig. 2) because it is confined to the southern part of eastern Europe, from Italy to Ukraine.

However, the populations of the *central* and *peripheral* lineages did not expand until 5000 years later, after the beginning of the Holocene interglacial. A similar situation has previously been found in the field vole *M. agrestis* (Linneaus, 1761), another Eurasian rodent, where the six extant cyt *b* lineages appear to have expanded at the beginning of the Holocene (Herman & Searle, 2011). These field vole lineages are assumed to have originated within geographically dispersed refugia at the time of the Younger Dryas (Herman & Searle, 2011), the cold period which separated the Bølling-Allerød from the Holocene (approximately 11.7 to 12.9 kya; Steffensen *et al.*, 2008). Although the *central* and *peripheral* lineages of the wood mouse originated at an earlier time, around the LGM (approximately 21.0 kya), their populations may also have gone through bottlenecks at the Younger Dryas climatic minimum, in which case these single locus skyline plots would only recover the most recent expansion subsequent to this event. The partial replacement of the *peripheral* lineage by the *central* lineage would then have occurred during the Holocene expansion of the two lineages after the Younger Dryas.

This scenario can be contrasted with that based on the *Apodemus* clock rate:

The *Apodemus* clock rate (Suzuki *et al.*, 2015) is approximately 2.3 times lower than the *Microtus* rate that we used, and would have led to time estimates increased by that factor. The resulting tMRCAs for the three mainland European lineages (approximately 38–52 ka) might conceivably relate to population bottlenecks at earlier temperature minima within the Weichselian glacial period (Johnsen *et al.*, 2001). Although the subsequent demographic expansions of the *central* and *peripheral* lineages would then broadly coincide with the post-glacial climatic warming after the LGM, the somewhat earlier expansion of the *south-eastern* lineage would be aligned with the coldest period towards the end of the Weichselian (Johnsen *et al.*, 2001) and there would be no post-glacial signal of expansion in this lineage. The scenario based on the *Apodemus* clock rate fits poorly with expectations in terms of population expansions, and we consider that the scenario based on the *Microtus* rate is much more realistic and so we continue to follow that here.

Although the overall phylogeographical pattern suggests that the mainland European wood mouse population is composed of three maternal lineages that colonized the continent from separate glacial refugia, it is nevertheless difficult to associate these putative refugia with any specific locations. Fossil remains of the wood mouse have been recorded from LGM sites in the eastern Pyrenees, northern Spain, the Dordogne region of southern central France, and from north-eastern Italy, to the south-east of the Alps (Sommer & Nadachowski, 2006). In addition, there are sporadic records of wood mouse from the Carpathian basin around the time of the LGM, although the species does not make a substantial contribution to the fauna there until the Holocene (Pazonyi, 2004). Only one of these refugia, in north-eastern Italy, can be placed within the exclusive range of a lineage, the *south-eastern*. The eastern Pyrenees are a possible source of the *central* lineage, given their location on the edge of the Iberian peninsula, while the Dordogne or the Carpathians might be the source of the *peripheral* lineage. Nevertheless, it is interesting that the wood mouse has only rarely been found in Carpathian sites from the LGM, from where other temperate species have been inferred to have colonized Europe after the last glaciation (Kotlík *et al.*, 2006; Wójcik *et al.*, 2010; McDevitt *et al.*, 2012; Stojak *et al.*, 2015). Tentative support for the Dordogne as the refugium for the *peripheral* lineage is provided by species distributional modelling (Fløjgaard *et al.*, 2009) because their reconstruction of suitable LGM habitat for the wood mouse included the Mediterranean coast of France, which is quite close to the Dordogne but did not include the Carpathian basin. According to the species distribution model, potentially suitable LGM habitat was also present in the Iberian peninsula, now occupied only by animals from the *central*

lineage, and in the current exclusive range of the *south-eastern* lineage.

The genetic divergence of wood mice from Sicily has previously been noted (Michaux *et al.*, 1998, 2003) and the *Sicilian* lineage does indeed appear to be older than the others, based on both its mitochondrial genetic diversity and time of coalescence (approximately 29 kya). The delay in the onset of its demographic expansion, until approximately 8000 years later, might conceivably be a result of the limited availability of suitable habitat on the island before the climate began to warm after the LGM. It is much earlier than the expansions of the other lineages, although this is not surprising, given its Mediterranean location, whereas the expansion of the other lineages would have been constrained until later by the climatic conditions further north in the continent. Application of the clock rate inferred for *Apodemus* (Suzuki *et al.*, 2015) would place the tMRCA of the *Sicilian* lineage at approximately 68 kya, which coincides with temperature minima around Marine Isotope Stage (MIS) 4 (Johnsen *et al.*, 2001) and is therefore plausible. However, the gradual rise in the population size (Fig. 4) would have lasted from approximately 25 to 50 kya, within the Weichselian glacial period.

The timescale for closure of the Messina Strait, based on sophisticated geophysical models, demonstrates the existence of a land-bridge between Sicily and the Italian mainland from approximately 17 to 27 Kya (Antonioli *et al.*, 2012, 2014), With the radiocarbon dated *Microtus* clock rate, the median tMRCA of the *Sicilian* lineage is about two thousand years earlier than the appearance of the land-bridge, although the connection was present within the 95% highest posterior density (HPD) interval of the tMRCA (19.6 to 40.5 kya). With the clock rate derived from *Apodemus*, the most recent 95% confidence limit for the tMRCA of the Sicilian lineage was approximately 46 kya, which is well before the time of the land bridge. However, there is fossil evidence of wood mice from Sicily that has been dated at approximately 32 $\pm$ 4 kya (Bonfiglio *et al.*, 2008), prior to the presence of the land-bridge, and it therefore appears that wood mice, although not a part of the original endemic island fauna, were able to reach the island with the early influx of mainland species in the late glacial (Bonfiglio *et al.*, 2001, 2002, 2008).

Although the isolated wood mouse population of Sicily appears to have arrived by some other means, the *African* and *Channel Island* lineages were most likely introduced accidentally to these outlying locations by human agents, following their original overland colonization of mainland Europe. These two lineages are much more recent in origin than the mainland or *Sicilian* ones, with median tMRCA estimate of approximately 7.4 kya and 5.6 Kya respectively (Table 3), based on the radiocarbon dated *Microtus* molecular clock rate (Martínková *et al.*, 2013). The *African* lineage is closely related to the *central* lineage (Fig. 3), the only one found in the Iberian peninsula (Fig. 2), suggesting that the wood mouse reached North Africa from there, rather than elsewhere in southern Europe. The demographic expansion of the *African* lineage followed closely after its origin (Fig. 4). Furthermore, the relative timing of expansions in different North African locations, inferred from mismatch analyses of cyt *b* sequences, provide evidence of an earlier introduction to Morocco than Algeria, suggesting that the wood mouse colonized the continent via the Strait of Gibraltar (Lalis *et al.*, 2016). The dates of the earliest wood mouse fossils from the Maghreb, recovered in the Tingitana Peninsula of Morocco (6–7.5 kya) and in Algeria (2.5–4 kya) fit well with these relative times of expansion (Lalis *et al.*, 2016; *sensu* Stoetzel, 2009, 2013). However, with the clock rate inferred for *Apodemus* (Suzuki *et al.*, 2015), the tMRCA of the African lineage would be approximately 17 Kya, indicating that the species colonized the continent immediately after the last glaciation, and pre-dating the fossil evidence.

The close relationship between the *African* and *Channel Island* lineages can most reasonably be explained by an introduction from one location to the other, most likely from Africa to the Channel Islands, or introduction to their present locations from a common source. If the latter, there is no precise indication of where this might be, although it would most likely be located within the Iberian peninsula. With the radiocarbon dated clock rate, the inferred timing of the introductions to the Channel Islands (approximately 5.6 Kya; median tMRCA) suggests that the mice may have been translocated by Neolithic people but, given the 95% HPD interval of the tMRCA (1.3–11.5 kya), later introductions are also possible. Interestingly, introductions around 3 kya between the Mediterranean and north-western Europe (Britain, northern France and nearby areas) have been suggested for the house mouse (Jones *et al.*, 2013). Using the *Apodemus* clock rate (Suzuki *et al.*, 2015), the wood mouse would have been introduced to the Channel Islands approximately 11.5 kya, before the Neolithic.

Although most of the genetic variation was partitioned among the six lineages, a substantial proportion (23%) was a result of the differences among the sequences within them. However, despite their overall variability, little geographical structure was discernible within the lineages and closely-related sequences were frequently obtained from widely separated localities, sometimes at opposite ends of the

species' range. This pattern is remarkable for a small mammal, once again suggesting that wood mice have a high capacity for dispersal. Despite the lack of geographical structure within lineages, it is possible to gain some information about the pattern of colonization from the presence of specific haplotypes at more than one location. In most cases, this could be attributed to natural overland colonization, although this might (in some cases) have been mediated by humans. The movement of animals between some shared locations would involve water crossings (see Supporting information, Fig. S3), sometimes between islands that have been separated throughout the post-glacial period (Sturt *et al.*, 2013). Dispersal to or from these islands must surely be a result of human activity, and so sharing of haplotypes between these locations provides a means to examine the effect of human influence, overlaid upon the otherwise natural distribution of the species.

All of the inferred translocations relating to presence of specific haplotypes in more than one location and definitively involving water crossings were between the British Isles, Iceland and the southern part of the Fennoscandian Peninsula (see Supporting information, Fig. S3), which might implicate Norse Viking settlers in the process. This has previously been proposed for the wood mouse (Berry *et al.*, 1967; Berry, 1969, 1973) and for the house mouse (Searle *et al.*, 2009a; Jones *et al.*, 2012, 2013). A similar hypothesis has also been put forward to explain the presence of Scandinavian mtDNA haplotypes in Irish badgers *Meles meles* (Frantz *et al.*, 2014). Wood mice probably reached Iceland with Norse people because historical documentation indicates that they were the first people to colonize the island, approximately 1100 years BP (Smith, 1995; Price & Gestsdóttir, 2006). However, as the islands around Britain have been occupied since the Mesolithic (Corbet, 1961), animals could have been translocated at any time since then, and this is reflected in the presence of the *peripheral* lineage throughout the British Isles. The lack of observed mitochondrial genetic structure across the wide ranges of the *central* and *peripheral* lineages, and the presence of the wood mouse on so many islands (Montgomery, 1999b), together show that the species has a high capacity for dispersal by human agency. It therefore appears likely that wood mice have been transferred repeatedly between locations throughout the Holocene, in what is probably an ongoing process. Interestingly, it is the subsequent translocations between the British Isles and Scandinavia, identified here through mitochondrial haplotype sharing, rather than the background of common ancestry between all of the *peripheral* lineage, that was identified in the earlier work based on non-metrical skeletal characters (Berry *et al.*, 1967; Berry, 1969, 1973).

The phylogeography of the wood mouse provides an interesting contrast with that of the house mouse. It appears to have survived the LGM in three mainland European locations, together with another population on the island of Sicily. The initial colonization of the mainland European continent by the wood mouse was apparently the result of a natural post-glacial expansion into newly available habitat and this was followed by a phase of further, human-mediated, introductions to numerous islands and otherwise inaccessible locations around the western and southern margins of the continent. The house mouse was a later arrival in Europe, introduced from its natural range in the east in conjunction with the gradual spread of human agriculture, conurbations, and trade (Bonhomme *et al.*, 2010). Although it too has been introduced to many islands on the margins of Europe, originally by Norse or Danish Vikings in the case of those around the Atlantic coasts, it has subsequently been carried all over the world as a true commensal (Jones *et al.*, 2013). The wood mouse, despite its capacity for human dispersal, has progressed little further than the margins of its natural range, presumably because it is anthropophilic, rather than a true commensal like the house mouse (Hulme-Beaman *et al.*, 2016).

## ACKNOWLEDGEMENTS

We thank the three anonymous reviewers for their helpful comments. We also thank the individuals and organizations who provided samples or advice: Paulo Alves, J. Ballantyne, David Bates, Sam Berry, Tom Black, Bristol City Museum and Art Gallery, Joan Carter, Norma Chapman, John Chester, Mike Cockram, Countryside Council for Wales, Shirley Cross, Charles David, Tim Deans, Fraser Dodds, Tim Dodman, K. Fairclough, Clem Fisher, Alain Frantz, Nick Gould, Great North Museum: Hancock, İslam Gündüz, Hampshire Museums, Gemma Harding, Mary Harman, Dawn Hayden, Alice Helyar, Herbert Art Gallery and Museum (Coventry), Martin Heubeck, Stephen Hewitt, Lister Hogarth, Phil Howard, Hugh Insley, Inverness Museum and Art Gallery, Maarit Jaarola, Jóhannes B. Jóhannsson, Gareth Jones, Andrew Kitchener, Guðmundur Ó. Kristjánsson, Steve Lane, Linley C. Lewis, Ian Linn, Liverpool Museum, Patrick Lowe, Herbert Mackenzie, Anne MacLellan, John Allan MacLellan, Natalia Martínková, Peter Mayhew, Damien McDevitt, Bob McGowan, Yvonne Meyer-Lucht, Alina Mishta, Eric Morton, Museum nationale d'Histoire naturelle

(Paris), National Trust for Scotland, North Lincolnshire Museum, Alan Ogden, Geoff Oxford, Helga Óskarsdóttir, Marine Pascal, Michel Pascal, Abbie Patterson, Joana Paupério, Les Pearce, Josephine Pemberton, Brian Rabbitts, Ramugondo V. Rambau, F. Ratter, Ian Ross, Royal Society for the Protection of Birds, Hazel Ryan, Mike Ryan, Kate Sampson, R. Sandling, Sue Scoggins, Scottish Natural Heritage, Mark Shaw, Ingibjörg Sigurjónsdóttir, Karl Skirnisson, R. Swann, Christine Taylor, Sandra Telfer, Nina Thomson, Steve Thomson, George Trafford, Sam Trebilcock, Tullie House Museum, R. Tulloch, Karen Varnham, Rodrigo Vega, Phil Wheeler, Sian Whitehead, Derek Yalden, and Grace Yoxon. We thank Barbara Marczuk and Iwona Ruczyńska for their help in the laboratory. Fieldwork in Iceland was supported by the Genetics Society Heredity Grant.

## REFERENCES

**Amori G, Hutterer R, Kryštufek B, Yigit N, Mitsain G, Palomo LJ. 2008.** *Apodemus flavicollis*. *The IUCN Red List of Threatened Species* **2008:**e.T1892A8699693.

**Anisimova M, Gascuel O. 2006.** Approximate likelihood-ratio test for branches: a fast, accurate, and powerful alternative. *Systematic Biology* **55:** 539–552.

**Anisimova M, Gil M, Dufayard J-F, Dessimoz C, Gascuel O. 2011.** Survey of branch support methods demonstrates accuracy, power, and robustness of fast likelihood-based approximation schemes. *Systematic Biology* **60:** 685–699.

**Antonioli F, Presti VL, Morticelli MG, Mannino MA, Lambeck K, Ferranti L, Bonfiglio L, Mangano G, Sannino GM, Furlani S, Sulli A, Palombo MR, Canese SP. 2012.** The land bridge between Europe and Sicily over the past 40 kyrs: timing of emersion and implications for the migration of *Homo sapiens*. *Rendiconti Online della Società Geologica Italiana* **21:** 1167–1169.

**Antonioli F, Presti VL, Morticelli MG, Bonfiglio L, Mannino MA, Palombo MR, Sannino G, Ferranti L, Furlani S, Lambeck K, Canese S, Catalano R, Chiocci FL, Mangano G, Scicchitano G, Tonielli R. 2014.** Timing of the emergence of the Europe-Sicily bridge (40-17 cal ka BP) and its implications for the spread of modern humans. In: Harff J, Bailey G, Lüth F, eds. *Geology and archaeology: submerged landscapes of the continental shelf*. *Geological Society, London, Special Publications* **411:** 111–144.

**Bandelt H, Forster P, Röhl A. 1999.** Median-joining networks for inferring intraspecific phylogenies. *Molecular Biology and Evolution* **16:** 37–48.

**Berry RJ. 1969.** History in the evolution of *Apodemus sylvaticus* (Mammalia) at one edge of its range. *Journal of Zoology* **159:** 311–328.

**Berry RJ. 1973.** Chance and change in British long-tailed field mice (*Apodemus sylvaticus*). *Journal of Zoology* **170:** 351–366.

**Berry RJ, Evans IM, Sennitt BFC. 1967.** The relationships and ecology of *Apodemus sylvaticus* from the Small Isles of the Inner Hebrides, Scotland. *Journal of Zoology* **152:** 333–346.

**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.

**Björck S. 1995a.** A review of the history of the Baltic Sea, 13.0–8.0 ka BP. *Quaternary International* **27:** 19–40.

**Björck S. 1995b.** Late Weichselian to early Holocene development of the Baltic Sea – with implications for coastal settlements in the southern Baltic region. In: Fischer A, ed. *Man and sea in the Mesolithic. Coastal settlement above and below present sea level. Proceedings of the International Symposium, Kalundberg, Denmark, 1993. Oxbow Monograph 53*. Oxford: Oxbow Books, 23–34.

**Bonfiglio L, Mangano G, Marra AC, Masini F. 2001.** A new late Pleistocene vertebrate faunal complex from Sicily (S. Teodoro Cave, North-Eastern Sicily, Italy). *Bollettino della Società Paleontologica Italiana* **40:** 149–158.

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

**Bonfiglio L, Esu D, Mangano G, Masini F, Petruso D, Soligo M, Tuccimei P. 2008.** Late Pleistocene vertebrate-bearing deposits at San Teodoro Cave (North-Eastern Sicily): preliminary data on faunal diversification and chronology. *Quaternary International* **190:** 26–37.

**Bonhomme F, Orth A, Cucchi T, Rajabi-Maham H, Catalan J, Boursot P, Auffray J-C, Britton-Davidian J. 2010.** Genetic differentiation of the house mouse around the Mediterranean basin: matrilineal footprints of early and late colonization. *Proceedings of the Royal Society of London Series B, Biological Sciences* **278:**1034–1043.

**Bouckaert R, Heled J, Kühnert D, Vaughan T, Wu C-H, Xie D, Suchard MA, Rambaut A, Drummond AJ. 2014.** BEAST 2: a software platform for Bayesian evolutionary analysis. *PLoS Computational Biology* **10:** e1003537.

**Burridge CP, Craw D, Fletcher D, Waters JM. 2008.** Geological dates and molecular rates: fish DNA sheds light on time dependency. *Molecular Biology and Evolution* **25:** 624–633.

**Conroy CJ, Cook JA. 1999.** MtDNA evidence for repeated pulses of speciation within arvicoline and murid rodents. *Journal of Mammalian Evolution* **6:** 221–245.

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

**Corbet GB. 1978.** *The mammals of the Palaearctic region: a taxonomic review*. London: British Museum (Natural History).

**Darriba D, Taboada GL, Doallo R, Posada D. 2012.** jModelTest 2: more models, new heuristics and parallel computing. *Nature Methods* **9:** 772.

**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.

**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.

**Dubey S, Michaux J, Brünner H, Hutterer R, Vogel P. 2009.** False phylogenies on wood mice due to cryptic cytochrome-*b* pseudogene. *Molecular Phylogenetics and Evolution* **50:** 633–641.

**Excoffier L, Lischer HE. 2010.** Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Molecular Ecology Resources* **10:** 564–567.

**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.

**Flowerdew JR, Tattersall FH. 2008.** Wood mouse *Apodemus sylvaticus*. In: Harris S, Yalden DW, eds. *Mammals of the British Isles: handbook*, 4th edn. Southampton: The Mammal Society, 125–137.

**Frantz AC, McDevitt AD, Pope LC, Kochan J, Davison J, Clements CF, Elmeros M, Molina-Vacas G, Ruiz-Gonzalez A, Balestrieri A, Van Den Berge K, Breyne P, Do Linh San E, Argen EO, Suchentrunk F, Schley L, Kowalczyk R, Kostka BI, Cirovic D, Sprem N, Colyn M, Ghirardi M, Racheva V, Braun C, Oliveira R, Lanszki J, Stubbe A, Stubbe M, Stier N, Burke T. 2014.** Revisiting the phylogeography and demography of European badgers (*Meles meles*) based on broad sampling, multiple markers and simulations. *Heredity* **113:** 443–453.

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

**Gouy M, Guindon S, Gascuel O. 2010.** SeaView version 4: a multiplatform graphical user interface for sequence alignment and phylogenetic tree building. *Molecular Biology and Evolution* **27:** 221–224.

**Guindon S, Dufayard JF, Lefort V, Anisimova M, Hordijk W, Gascuel O. 2010.** New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Systematic Biology* **59:** 307–321.

**Herman JS, Searle JB. 2011.** Post-glacial partitioning of mitochondrial genetic variation in the field vole. *Proceedings of the Royal Society of London Series B, Biological Sciences* **278:**3601–3607.

**Herman JS, McDevitt AD, Kawałko A, Jaarola M, Wójcik JM, Searle JB. 2014.** Land-bridge calibration of molecular clocks and the post-glacial colonization of Scandinavia by the Eurasian field vole *Microtus agrestis*. *PLoS ONE* **9:** e103949.

**Hewitt GM. 1999.** Post-glacial re-colonization of European biota. *Biological Journal of the Linnean Society* **68:**87–112.

**Hijmans RJ. 2015.** *geosphere: spherical trigonometry. R package, Version 1.4–3*. Available at: http://CRAN.R-project.org/package=geosphere.

**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, Kolokotronis S-O, Allaby RG. 2007.** Elevated substitution rates estimated from ancient DNA sequences. *Biology Letters* **3:** 702–705.

**Ho SYW, Lanfear R, Bromham L, Phillips MJ, Soubrier J, Rodrigo AG, Cooper A. 2011.** Time-dependent rates of molecular evolution. *Molecular Ecology* **20:** 3087–3101.

**Ho SYW, Duchene S, Molak M, Shapiro B. 2015.** Time-dependent estimates of molecular evolutionary rates: evidence and causes. *Molecular Ecology* **24:** 6007–6012.

**Hoofer SR, Gaschak S, Dunina-Barkovskaya Y, Makluk J, Meeks HN, Wickliffe JK, Baker RJ. 2007.** New information for systematics, taxonomy, and phylogeography of the rodent genus *Apodemus* (*Sylvaemus*) in Ukraine. *Journal of Mammalogy* **88:** 330–342.

**Hulme-Beaman A, Dobney K, Cucci T, Searle JB. 2016.** An ecological and evolutionary framework for commensalism in anthropogenic environments. *Trends in Ecology and Evolution* **31:** 633–645.

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

**Johnsen SJ, Dahl-Jensen D, Gundestrup N, Steffensen JP, Clausen HB, Miller H, Masson-Delmotte V, Sveinbjörnsdottir AE, White J. 2001.** Oxygen isotope and palaeotemperature records from six Greenland ice-core stations: Camp Century, Dye-3, GRIP, GISP2, Renland and NorthGRIP. *Journal of Quaternary Science* **16:** 299–307.

**Jones EP, Skirnisson K, McGovern TH, Gilbert MTP, Willerslev E, Searle JB. 2012.** Fellow travellers: a concordance of colonization patterns between mice and men in the North Atlantic region. *BMC Evolutionary Biology* **12:** 35.

**Jones EP, Eager HM, Gabriel SI, Jóhannesdóttir F, Searle JB. 2013.** Genetic tracking of mice and other bioproxies to infer human history. *Trends in Genetics* **29:** 298–308.

**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.

**Kotlík P, Marková S, Vojtek L, Stratil A, Šlechta V, Hyršl P, Searle JB. 2014.** Adaptive phylogeography: functional divergence between haemoglobins derived from different glacial refugia in the bank vole. *Proceedings of the Royal Society of London Series B, Biological Sciences* **281:**20140021.

**Lalis A, Leblois R, Liefried S, Ouarour A, Beeravolu CR, Michaux J, Hamani A, Denys C, Nicolas V. 2016.** New molecular data favour an anthropogenic introduction

of the wood mouse (*Apodemus sylvaticus*) in North Africa. *Journal of Zoological Systematics and Evolutionary Research*. **54:** 1–12.

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

**Macholán M, Mrkvicová Vyskočilová M, Bejček V, Šťastný K. 2012.** Mitochondrial DNA sequence variation and evolution of Old World house mice (*Mus musculus*). *Folia Zoologica* **61:** 284–307.

**Martin Y, Gerlach G, Schlötterer C, Meyer A. 2000.** Molecular phylogeny of European muroid rodents based on complete cytochrome *b* sequences. *Molecular Phylogenetics and Evolution* **16:** 37–47.

**Martínková N, Barnett R, Cucchi T, Struchen R, Pascal M, Pascal M, Fischer MC, Higham T, Brace S, Ho SYW, Quéré J-P, O'Higgins P, Excoffier L, Heckel G, Hoelzel AR, Dobney KM, Searle JB. 2013.** Divergent evolutionary processes associated with colonization of offshore islands. *Molecular Ecology* **22:** 5205–5220.

**McDevitt AD, Zub K, Kawałko A, Oliver MK, Herman JS, Wójcik JM. 2012.** Climate and refugial origin influence the mitochondrial lineage distribution of weasels *Mustela nivalis* in a phylogeographic suture zone. *Biological Journal of the Linnean Society* **106:** 57–69.

**Michaux JR, Sara M, Libois R, Matagne R. 1998.** Is the woodmouse (*Apodemus sylvaticus*) of Sicily a distinct species? *Belgian Journal of Zoology* **128:** 211–214.

**Michaux JR, Kinet S, Filippucci MG, Libois R, Besnard A, Catzeflis F. 2001.** Molecular identification of three sympatric species of wood mice (*Apodemus sylvaticus*, *A. flavicollis*, *A. alpicola*) in western Europe (Muridae: Rodentia). *Molecular Ecology Notes* **1:** 260–263.

**Michaux JR, Chevret P, Filippucci M-G, Macholan M. 2002.** Phylogeny of the genus *Apodemus* with a special emphasis on the subgenus *Sylvaemus* using the nuclear IRBP gene and two mitochondrial markers: cytochrome *b* and 12S rRNA. *Molecular Phylogenetics and Evolution* **23:** 123–136.

**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 JR, Libois R, Filippucci M-G. 2005.** So close and so different: comparative phylogeography of two small mammal species, the yellow-necked fieldmouse (*Apodemus flavicollis*) and the woodmouse (*Apodemus sylvaticus*) in the Western Palearctic region. *Heredity* **94:** 52–63.

**Mix AC, Bard E, Schneider R. 2001.** Environmental processes of the ice age: land, oceans, glaciers (EPILOG). *Quaternary Science Reviews* **20:** 627–657.

**Montgomery WI. 1999a.** *Apodemus flavicollis* (Melchior, 1834). In: Mitchell-Jones AJ, Amori G, Bogdanowicz W, Kryštufek B, Reijnders PJH, Spitzenberger F, Stubbe M, Thissen JBM, Vohralík V, Zima J, eds. *The atlas of European mammals*. London: T & AD Poyser, 270–271.

**Montgomery WI. 1999b.** *Apodemus sylvaticus* (Linnaeus, 1758). In: Mitchell-Jones AJ, Amori G, Bogdanowicz W, Kryštufek B, Reijnders PJH, Spitzenberger F, Stubbe M, Thissen JBM, Vohralik V, Zima J, eds. *The atlas of European mammals*. London: T & AD Poyser, 274–275.

**Musser GG, Carleton MD. 2005.** Superfamily Muroidea. In: Wilson DE, Reeder DM, eds. *Mammal species of the world. A taxonomic and geographic reference*, 3rd edn. Baltimore, MD: John Hopkins University Press, 894–1531.

**Nieberding C, Morand S, Libois R, Michaux JR. 2004.** A parasite reveals cryptic phylogeographic history of its host. *Proceedings of the Royal Society of London Series B, Biological Sciences* **271:**2559–2568.

**Nieberding C, Libois R, Douady CJ, Morand S, Michaux JR. 2005.** Phylogeography of a nematode (*Heligmosomoides polygyrus*) in the western Palearctic region: persistence of northern cryptic populations during ice ages? *Molecular Ecology* **14:** 765–779.

**Pazonyi P. 2004.** Mammalian ecosystem dynamics in the Carpathian Basin during the last 27 000 years. *Palaeogeography, Palaeoclimatology, Palaeoecology* **212:** 295–314.

**Piertney SB, Stewart WA, Lambin X, Telfer S, Aars J, Dallas JF. 2005.** Phylogeographic structure and postglacial evolutionary history of water voles (*Arvicola terrestris*) in the United Kingdom. *Molecular Ecology* **14:** 1435–1444.

**Posada D, Crandall KA. 2001.** Intraspecific gene genealogies: trees grafting into networks. *TRENDS in Ecology & Evolution* **16:** 37–45.

**Price TD, Gestsdóttir H. 2006.** The first settlers of Iceland: an isotopic approach to colonisation. *Antiquity* **80:** 130–144.

**R Core Team. 2015.** *R: a language and environment for statistical computing*. Vienna: R Foundation for Statistical Computing. Available at: https://www.R-project.org/

**Rajabi-Maham H, Orth A, Bonhomme F. 2008.** Phylogeography and postglacial expansion of *Mus musculus domesticus* inferred from mitochondrial DNA coalescent, from Iran to Europe. *Molecular Ecology* **17:** 627–641.

**Rambaut A, Suchard MA, Xie D, Drummond AJ. 2014.** *Tracer, Version 1.6*. Available at: http://beast.bio.ed.ac.uk/Tracer

**Reutter BA, Petit E, Brünner H, Vogel P. 2003.** Cytochrome *b* haplotype divergences in West European *Apodemus*. *Mammalian Biology* **68:** 153–164.

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

**Schlitter D, van der Straeten E, Amori G, Hutterer R, Kryštufek B, Yigit N, Mitsain G. 2008.** *Apodemus sylvaticus*. *The IUCN Red List of Threatened Species* **2008:** e.T1904A8791394.

**Schneider S, 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.

**Searle JB, Wilkinson PJ. 1987.** Karyotypic variation in the common shrew (*Sorex araneus*) in Britain - a 'Celtic Fringe'. *Heredity* **59:** 345–351.

**Searle JB, Jones CS, Gündüz İ, Scascitelli M, Jones EP, Herman JS, Rambau RV, Noble LR, Berry RJ, Giménez MD, Jóhannesdóttir F. 2009a.** Of mice and

(Viking?) men: phylogeography of British and Irish house mice. *Proceedings of the Royal Society of London Series B, Biological Sciences* **276:**201–207.

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

**Shimodaira H, Hasegawa M. 1999.** Multiple comparisons of log-likelihoods with applications to phylogenetic inference. *Molecular Biology and Evolution* **16:** 1114–1116.

**Smith KP. 1995.** Landnám: the settlement of Iceland in archaeological and historical perspective. *World Archaeology* **26:** 319–347.

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

**Steffensen JP, Andersen KK, Bigler M, Clausen HB, Dahl-Jensen D, Fischer H, Goto-Azuma K, Hansson M, Johnsen SJ, Jouzel J, Masson-Delmotte V, Popp T, Rasmussen SO, Röthlisberger R, Ruth U, Stauffer B, Siggaard-Andersen M-L, Sveinbjörnsdóttir ÁE, Svensson A, White JWC. 2008.** High-resolution Greenland ice core data show abrupt climate change happens in few years. *Science* **321:** 680–684.

**Stoetzel E. 2009.** Les Microvertébrés du site d'occupation humaine d'El Harhoura 2 (Pléistocène Supérieur-Holocène, Maroc): Systématique, Évolution, Taphonomie et Paléoécologie. PhD Thesis, Paris: Muséum national d'Histoire naturelle.

**Stoetzel E. 2013.** Late Cenozoic micromammal biochronology of northwestern Africa. *Palaeogeography, Palaeoclimatology, Palaeoecology* **392:** 359–381.

**Stojak J, McDevitt AD, Herman JS, Searle JB, Wójcik JM. 2015.** Post-glacial colonization of eastern Europe from the Carpathian refugium: evidence from mitochondrial DNA of the common vole *Microtus arvalis*. *Biological Journal of the Linnean Society* **115:** 927–939.

**Sturt F, Garrow D, Bradley S. 2013.** New models of North West European Holocene palaeogeography and inundation. *Journal of Archaeological Science* **40:** 3963–3976.

**Suzuki H, Filippucci MG, Chelomina GN, Sato JJ, Serizawa K, Nevo E. 2008.** A biogeographic view of *Apodemus* in Asia and Europe inferred from nuclear and mitochondrial gene sequences. *Biochemical Genetics* **46:** 329–346.

**Suzuki Y, Tomozawa M, Koizumi Y, Tsuchiya K, Suzuki H. 2015.** Estimating the molecular evolutionary rates of mitochondrial genes referring to Quaternary ice age events with inferred population expansions and dispersals in Japanese *Apodemus*. *BMC Evolutionary Biology* **15:** 187.

**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.

**Tamura K, Nei M. 1993.** Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. *Molecular Biology and Evolution* **10:** 512–526.

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

**Wójcik JM, Kawałko 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.

## SUPPORTING INFORMATION

Additional Supporting Information may be found online in the supporting information tab for this article:

**Figure S1.** Maximum likelihood (ML) tree for wood mouse sequences (monophyletic clade collapsed), together with four *Apodemus* outgroup and four wood mouse *numt* sequences (GenBank reference numbers shown). Branch supports are SH-aLRT obtained with PHYML, version 3.0.

**Figure S2.** Mismatch distributions for six wood mouse cytochrome *b* (cyt *b*) lineages. Observed (solid line) and expected (dashed line) distributions of pairwise differences between sequences within lineages.

**Figure S3.** Connections between locations with the same haplotype. Lines represent connections between all pairs of locations with the same haplotype, where movements between such locations would involve a sea crossing, given the present location of coastlines. Lines coloured according to lineage (Fig. 2). Locations are collapsed to mean or centroid coordinates within each land mass or archipelago. Coordinates of locations and geodesic distances between them in supplementary information (Table S4) and coastline data from *OpenStreetMap* (http://www.openstreetmap.org).

**Table S1.** Polymerase chain reaction (P) and sequencing (S) primers used to obtain the new cytochrome *b* (cyt *b*) sequences reported in the present study, from frozen or ethanol (EtOH) preserved tissues and from dry tissues or museum skins.

**Table S2.** Details of 981 cytochrome *b* (cyt *b*) sequences from *Apodemus sylvaticus*. Lineage, GenBank number, voucher (where available) or sample number, haplotype name, cyt *b* sequence length, locality, WGS84

coordinates and source of material. Sequences differing only by missing base calls treated as distinct haplotypes. Voucher codes refer to collections of the Mammal Research Institute, Polish Academy of Sciences, Białowieża (MRI.PAS) and National Museums of Scotland, Edinburgh (NMS.Z). *New sequences from the present study.

**Table S3.** Prior parameter distributions from coalescent genealogy sampling with BEAST, version 2.3.1. Linked substitution and clock model parameters, unlinked tree parameters.

**Table S4.** Locations with the same haplotype. Geographical coordinates for locations and pairwise geodesic distances between them, where movements between such locations would involve a sea crossing, given the present location of coastlines. Locations collapsed to mean or centroid coordinates within each land mass or archipelago.