*Article*

# Characterizing Mitochondrial Capture in an Iberian Shrew

**Henry D. Kunerth** [1], **Joaquim T. Tapisso** [2], **Raul Valente** [3, 4, 5] , **Maria da Luz Mathias** [2], **Paulo C. Alves** [3, 4, 6] , **Jeremy B. Searle** [1, 3, 4, *, †], **Rodrigo Vega** [1, 7, †] **and Joana Paupério** [3, 8, †]

[1]   Department of Ecology and Evolutionary Biology, Cornell University, Ithaca, NY 14853, USA; henrykunerth@gmail.com; https://orcid.org/0000-0002-2324-7773 (H.D.K.); jeremy.searle@cornell.edu; https://orcid.org/0000-0001-7710-5204 (J.B.S.)

[2]   CESAM – Centro de Estudos do Ambiente e do Mar, Departamento de Biologia Animal, Faculdade de Ciências, Universidade de Lisboa, Lisbon, Portugal; jstapisso@fc.ul.pt; https://orcid.org/0000-0002-8664-9956 (J.T.T.); mmathias@fc.ul.pt; https://orcid.org/0000-0003-3876-958X (M.L.M.)

[3]   CIBIO, Centro de Investigação em Biodiversidade e Recursos Genéticos, InBIO, Laboratório Associado/BIOPOLIS Program in Genomics, Biodiversity and Land Planning, Campus de Vairão, 4485-661 Vairão, Portugal; raulfonsecavalente@gmail.com; https://orcid.org/0000-0001-8544-373X (R.F.V.); pcalves@fc.up.pt; https://orcid.org/0000-0003-4797-0939 (P.C.A.); joanapcastro@cibio.up.pt; https://orcid.org/0000-0003-2569-0768 (J.P.)

[4]   Departamento de Biologia, Faculdade de Ciências, Universidade do Porto, Porto, Portugal

[5]   CIMAR/CIIMAR - Interdisciplinary Centre of Marine and Environmental Research, University of Porto, Avenida General Norton de Matos, S/N, 4450-208, Matosinhos, Portugal

[6]   Estação Biológica de Mértola (EBM), Praça Luís de Camões, 7750-329 Mértola, Portugal

[7]   Ecology Research Group, Section of Natural and Applied Sciences, School of Psychology and Life Sciences, Canterbury Christ Church University, Canterbury, UK; rodrigo.vega@canterbury.ac.uk; https://orcid.org/0000-0003-4915-9840

[8]   EMBL-EBI, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SD, UK

*   Correspondence: jeremy.searle@cornell.edu

† These authors contributed equally to this work.

**Abstract:** Mitochondrial introgression raises questions of biogeography, and extent of reproductive isolation and natural selection. Previous phylogenetic work on the *Sorex araneus* complex revealed apparent mitonuclear discordance in Iberian shrews indicating past hybridization of *S. granarius* and the Carlit chromosomal race of *S. araneus*, enabling introgression of the *S. araneus* mitochondrial genome into *S. granarius.* To further study this, we genetically typed 61 *S. araneus/coronatus/granarius* from localities in Portugal, Spain, France and Andorra at mitochondrial, autosomal and sex-linked loci, and combined our data with the previously published sequences. Our data are consistent with the earlier data that *S. coronatus* and *S. granarius* are the most closely related of the three species and confirm that *S. granarius* from the Central System mountain range in Spain captured the mitochondrial genome from a population of *S. araneus*. The mitochondrial capture event can be explained by invoking a biogeographical scenario whereby *S. araneus* was in contact with *S. granarius* during the Younger Dryas in central Iberia despite the two species currently having disjunct distributions. We discuss whether selection favoured *S. granarius* with an introgressed mitochondrial genome. Our data also suggest recent hybridization and introgression between *S. coronatus* and *S. granarius* and between *S. araneus* and *S. coronatus.*

**Keywords:** *Sorex araneus* complex; karyotype; introgression; phylogenetics; hybridization; Iberia

## 1. Introduction

Mitochondrial (mt) DNA sequences have long been exceptionally valuable molecular markers for systematics and specimen-based population genetics [1]. Even early in the population genetic study of mtDNA, instances of "cytonuclear discordance" [2] were demonstrated, including in small mammals [3], where populations are characterized by the nuclear genome (and morphology) of one species or subspecific taxon, but the mitochondrial genome of another species/subspecific taxon. This mitochondrial introgression, which has now been rather frequently observed, has been viewed as one

reason to be sceptical about use of mtDNA in phylogenetic and population genetic inference [4], but increasingly researchers are taking the view that instances of cytonuclear discordance are of interest in themselves in describing an aspect of population history. In particular, there is a realization that integration of a foreign mitochondrial genome may often, or perhaps typically, be advantageous, with selection favouring the process [5]. Hence the connotation associated with the term "mitochondrial capture" [e.g. 6] may be appropriate, and where mitochondrial introgression is observed it is worth thinking what selective process may be involved. Demographic processes and sex-biased dispersal may also be important [2]. It is an appreciation of what promotes mitonuclear discordance that adds nuance to our understanding of the recent population history of a species.

To gain an understanding of instances of mitochondrial capture requires widespread sampling over the distribution of a species or within an area of particular geographic interest. If possible, it is good to have widespread sampling of multiple species in the same genus, to help understand if it is the particular circumstance of the interaction of two particular species within the genus that leads to a capture event, or whether it is more pervasive than that. An exemplary case are hares of the genus *Lepus* in Europe and North America. Here mitochondrial capture has been demonstrated in multiple species, and in particular it is the mtDNA of more cold-adapted species that have introgressed into more warm-adapted species and these events occurred during periods of sympatry during glacial cold periods when the cold-adapted species have been widespread [e.g. 7-10].

The *Sorex araneus* complex consists of ten species of red-toothed shrews with a Holarctic distribution extending from western Europe to eastern North America, each distinguished by a different karyotype [11]. One of the species, *Sorex granarius*, an Iberian endemic, has a karyotype that may be ancestral to all the six European species [11,12]. It is this species that apparently has captured the mitochondrial genome of another species in the complex – *S. araneus* s.s. – according to studies with molecular markers by Yannic et al. [13,14]. Based on mitochondrial DNA, the phylogeny of members of the *S. araneus* complex in Iberia (*araneus*, *coronatus*, *granarius* [15]) show *S. granarius* most closely related to *S. araneus*, while phylogenies based on Y chromosome or autosomal markers show *S. granarius* most closely related to *Sorex coronatus*. As well as being distinguished from closely related species by karyotype, *S. araneus* are subdivided into over 70 karyotypically-distinct forms ("chromosomal races") [11]. Of these karyotypic forms, the one found in Iberia is named as the "Carlit" chromosomal race, which has a relatively undifferentiated karyotype – one of the most similar to the "ancestral" *S. granarius* karyotype [11]. Thus, the mitonuclear discordance found in Iberian shrews indicated to Yannic et al. [14] that there had been hybridization of *S. granarius* and the Carlit chromosomal race of *S. araneus*, enabling introgression of the *S. araneus* mitochondrial genome into *S. granarius*. A prediction from Yannic et al.'s hypothesis is that there is or was a "true" mitochondrial genome in *S. granarius*, this being the genome displaced by the *S. araneus* mitochondrial genome over at least part of the distribution range of *S. granarius*.

We here describe what we believe to be that predicted original *S. granarius* mitochondrial genome, identified in the context of a phylogenetic analysis of mtDNA sequences of the three Iberian species of the *S. araneus* complex. While there is a large dataset of *S. araneus* mtDNA sequences already available, this study expands the number of mtDNA sequences available for *S. granarius* and *S. coronatus* which enables a more thorough phylogenetic analysis of mtDNA variation in these latter species. Likewise, the new individuals were sequenced with nuclear markers to expand the analysis of mitonuclear discordance. We thus support the suggestion from Yannic et al. [14] of a mitochondrial capture event in Iberian shrews, and we consider how it might have come about. Our studies also provide preliminary evidence of other hybridization and introgression events among the three species of *Sorex* in Iberia.

## 2. Materials and Methods

### 2.1. Sampling, DNA extraction, and sequencing

DNA from 61 individuals within the *S. araneus* complex, specifically *S. araneus* s.s., *S. coronatus* and *S. granarius* was used in the study. These samples were collected between 1994 and 2016 across 23 localities in Portugal, Spain, France and Andorra (Figure 1). All tissues were stored in ethanol prior to extraction. DNA was extracted from foot, tail and ear tissues using Genomic DNA Minipreps Tissue Kits (EasySpin). DNA from five genes was amplified and sequenced. The loci included a mitochondrial locus *Cytochrome B* (*CytB*), a Y-linked locus *DBY7*, an X-linked locus *Zinc Finger Protein 1* (*ZFX1*) and two autosomal genes *Breast cancer susceptibility 1* (*BRCA1*) and *Apolipoprotein B* (*ApoB*).

A total of 2,970 base pairs (bp) of DNA was amplified and sequenced, 1,003 bp from *CytB*, 504 bp were from *DBY7*, 190 bp from *ZFX1*, 499 bp from *ApoB* and 774 bp from *BRCA1* (Supplementary Table S1). PCRs were performed in a final volume of 5 μL containing 0.5-1.0 μL of DNA, 2 μL of MasterMix, 0.2 μL of each primer and the remaining volume in water. Amplification was performed by 15 min initial denaturation at 95C followed by 35 cycles of 95C for 30 s (45 s for *CytB*), an annealing step for 30 s (60 s for *CytB* and 45 s for *ApoB* and *BRCA1*, see annealing temperatures in Supplementary Table S1), and 72C for 60 s (30 s for *ZFX1*) and a final extension step at 60C for 5 min. A touchdown protocol was applied to *DBY7*, with 21 cycles at 95C for 30 s, and decreasing annealing temperatures from 65C to 55C decreasing by 0.5C per cycle with otherwise identical parameters. PCR products were purified with ExoSAP-IT® PCR clean-up kits (GE Healthcare, Piscataway, NJ, USA). Sequencing reactions were performed using the BigDye® Terminator v1.1 Cycle Sequencing Kit (Applied Biosystems, Carlsbad, CA, USA). Samples were sequenced on both strands on a 3130xl Genetic Analyzer Sequencer (Applied Biosystems/HITASHI).

Sequences were assembled and edited in Sequencher v5.4.6 (Gene Codes Corp., Ann Arbor, MI, USA) and verified by eye, then aligned using ClustalW [16] implemented in MEGA v11 [17]. Additional sequences from Yannic et al. [13,14] from species *S. araneus*, *S. coronatus*, *S. granarius* and outgroup *Sorex alpinus* were also included and aligned. Samples were phased using the PHASE algorithm [18] implemented in DNAsp v6 [19]. All samples for each gene were phased using 100 iterations after 100 burn-in iterations.

### 2.2. Tree building

Maximum-likelihood trees were built using IQtree v1.6.12 [20] using the ModelFinder algorithm to automatically identify the best fit model for sequence evolution. Identical sequences were ignored for the purpose of tree building but presented in visualizations and subsequent analysis. Individual samples with heterozygous genotypes were included and each ambiguous character were given equal weight when building trees.

Haplotype networks were built in R using minimal spanning trees implemented through the R program Ape 5.6-2 [21].

### 2.3. Summary statistics

Diversity indices, including haplotype counts, haplotype diversity, nucleotide diversity, number of segregating sites, as well as the neutrality tests Tajima's *D* [22] and Fu's $F_s$ [23], were estimated using DNAsp v6 [19]. For each set of statistics, data was partitioned in multiple ways: including by geography, nuclear genomic species identity and mitochondrial species identity.

## 3. Results

Our data for one mitochondrial and four nuclear loci for 61 *S. araneus/coronatus/granarius* from Portugal, Spain, France and Andorra (Supplementary Table S2) were combined with data for the same loci on 35 individuals of the same three species from Yannic et al. [14] [*BRCA1*: GU473723-GU473750, GU473766-GU473780,

GU473786-GU473787; *ApoB*: GU473788-GU473816, GU473830-GU473840, GU473845-GU473846; *ZFX1*: GU473847-GU473870, GU473883-GU473893, GU473898-GU473899] supplemented with earlier published data [13] on *CytB* (EF636497-EF636516, EF636526) and *DBY7* (EF636575, EF636576, EF636580, EF636581, EF636585). The data from Yannic et al. were from Spain, France, Switzerland and Poland. The sampling points of all 96 individuals are mapped in Figure 1. We have followed Yannic et al. for the species designation of their specimens. For our specimens, we did not attempt to identify them morphologically (the species are very similar in appearance [14]). However, based on the geographical distribution of the three species provided by Yannic et al. [14], 52 of the specimens that we genotyped could be attributed unequivocally to one of the three species. Four specimens from northwestern Spain were from close to the contact area of *S. coronatus* and *S. granarius* while five specimens were from close to the contact area of *S. araneus* and *S. coronatus* in the vicinity of the Pyrenees, at the border of Spain and France (Supplementary Table S2). In our phylogenetic analyses we have attributed these nine individuals to species, but will justify that on the basis of haplotypes at the markers scored.

For *CytB*, there was a clear clade of *S. coronatus* haplotypes with subdivision into three subclades based on geography (Figure 2). All the French specimens were in one subclade together with specimens from nearby areas of Spain as far west as Burgos (black arrow in Figure 1; Supplementary Table S2). A second subclade consisted of specimens from Spain, from Burgos westwards. A third subclade consisted of specimens from Switzerland. The sister clade to the *coronatus* clade consisted largely of *S. araneus*, mostly in a single subclade, but a second smaller subclade included both *S. araneus* from the Pyrenees and *S. granarius* from the Central System mountain range in Spain. Even though one of the subclades had *CytB* deriving from two species, its nucleotide diversity (0.00379, $N$ = 10) was similar to that in other subclade (0.00305, $N$ = 21) and considerably lower than *S. araneus* overall (0.00608, $N$ = 27)(Supplementary Table S3). Finally, close to the outgroup branch in the tree there were eight distinct haplotypes of *S. granarius* from Portugal and northwestern Spain.

Haplotype diversity was very high (close to 1) and both Tajima's $D$ and Fu's $Fs$ were negative for all groupings in Supplementary Table S3, except for some of the groupings with small sample sizes. These are features of populations that have not been through extreme bottlenecks, but have undergone expansion relatively recently. The high diversity values are not surprising given that the specimens usually come from wide geographical areas.

The haplotypes of shrews collected in regions where two species occur are highlighted in Figure 2. The four individuals typed at *CytB* from the contact area of *S. coronatus* and *S. araneus* in the Pyrenees all had *CytB* haplotypes that clearly fit into one species or the other (three individuals as *S. araneus* [two haplotypes], one as *S. coronatus*), and in the appropriate subclades of those species given their geographic location. They also had nuclear genotypes consistent with their species designation (see below).

The two individuals typed at *CytB* from the contact area of *S. coronatus* and *S. granarius* were classified as *S. granarius*. There is no information on their nuclear genotype, however they have *CytB* haplotypes similar to *S. granarius* from Portugal, part of the same continuous distribution area (Figure 1). They are classified as *S. granarius* in this tree, but it is notable that the two other individuals from the same locality are classified as *S. coronatus* based on nuclear genotype (see below).

In the phylogenetic network based on *CytB* haplotypes, again *S. granarius* from Portugal and northwestern Spain was clearly positioned as the lineage closest to the outgroup, and the other *S. granarius* haplotypes were well distant on the network as part of the cluster of *S. araneus* haplotypes (Figure 3). As also evident in Figure 2, there was much more variation in *S. coronatus* than *S. araneus*. This is also clear from the diversity statistics, e.g. nucleotide diversity of 0.01371 ($N$ = 43) for *S. coronatus* and 0.00608 ($N$ = 27) for *S. araneus* (Supplementary Table S3). In the network, the *S. coronatus* were positioned in an intermediate position between *S. granarius* from western Iberia and the *S. araneus* together with *S. granarius* from the Central System mountain range (Figure 3).

The ML trees for the two autosomal markers, *ApoB* and *BRCA1* are shown in Figures 4 and 5, respectively. For neither marker was there a well-supported phylogenetic relationship between *S. granarius* and any *S. araneus*, contrary to what was found for *CytB*. For *BRCA1* the *S. granarius* from the Central System formed a well-supported clade together with *S. granarius* from Portugal distinct from all other *Sorex*. For *ApoB*, the *S. granarius* are more scattered in the tree but the Central System haplotype is identical to a haplotype found in *S. granarius* in Portugal, and all the *S. araneus* grouped together in a reasonably well-supported clade elsewhere in the tree.

For the Y-linked marker, *DBY7*, the ML tree showed a similar result as the autosomal markers, with the *S. granarius* from the Central System and Portugal forming a highly supported monophyletic group and all *S. araneus* also forming a clade (though not strongly supported) elsewhere in the tree (Figure 6). The X-linked marker *ZFX1* was the only one where *S. araneus* were not phylogenetically discrete from *S. granarius* although the support for the all the deeper nodes within the tree was very low (Figure 7).

The two individuals typed at nuclear markers from the contact area of *S. coronatus* and *S. granarius* in northwestern Spain were classified as *S. coronatus*, in contrast with the findings for the mitochondrial marker (see above; Supplementary Table 2). They were typed at two nuclear markers each (*DBY7* and *ZFX1*, and *ApoB* and *ZFX1*, respectively: Supplementary Table S2) and the positioning of the shrews in each case were clearly within the expectations for *S. coronatus* and distinct from those for *S. granarius* (Figures 4, 6 and 7). For the contact area of *S. coronatus* and *S. araneus* in the Pyrenees, all five individuals were typed at *ApoB* (Supplementary Table S2) and were present in the same species clade or grouping as expected on the basis of *CytB* haplotype (Supplementary Table S2; Figures 2, 4). There was one individual without a *CytB* haplotype whose identity as *S. araneus* is solely based on the *ApoB* result (SM.2182; Supplementary Table S2; Figure 4). Other nuclear markers supported the identifications of individuals from this contact area. However, there was an unexpected aspect to the results with *BRCA1* when considering a wider distribution area (Figure 5). Individuals otherwise classified as *S. coronatus* from northern Spain (in Cantabria [SM.CA.0340, 1187], Burgos [SM.BU.1193] and Picos de Europa [SM.AS.0323, SM.2175]) all had the same *BRCA1* haplotype as a *S. araneus* individual from within the contact area between *S. coronatus* and *S. araneus* in the Pyrenees (SM.2178)(Supplementary Table S2). Phylogenetically the *BRCA1* haplotype is clearly *S. araneus*.

Regarding the relationships of *Sorex araneus/coronatus/granarius*, in the *CytB* tree, *S. araneus* and *S. coronatus* were most closely related, while the network put the west Iberian *S. granarius* and the *S. coronatus* as most closely related. *S. granarius* and *S. coronatus* were also the most closely related species pair for *ApoB* and *DBY7*. For *BRCA1* and *ZFX1*, *S. araneus* and *S. granarius* appeared most closely related, but with very low bootstrap support in each case.
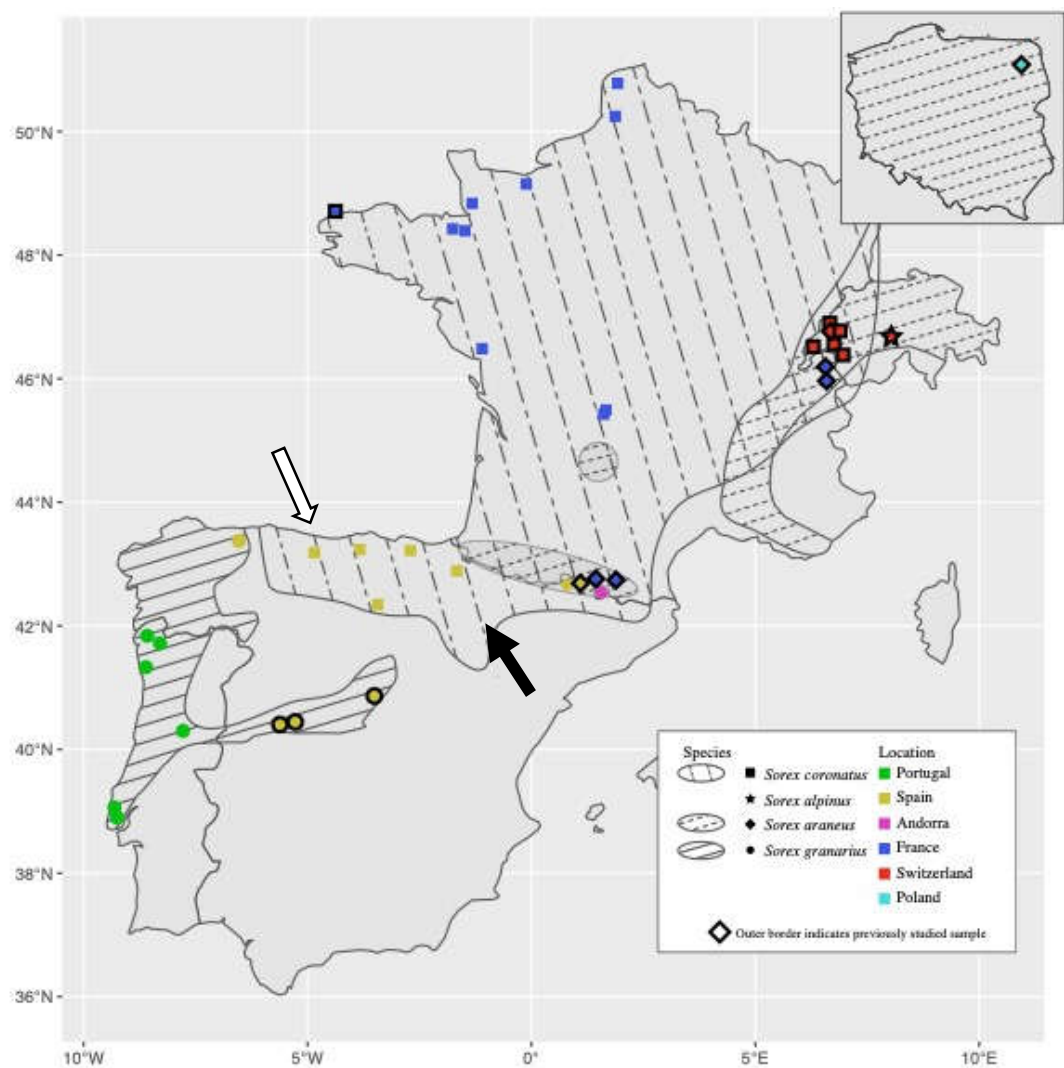
**Figure 1.** Collection localities of 61 *Sorex* shrews described here, plus 35 described by Yannic et al. [14] (see text). The specimens are classified as *S. araneus/coronatus/granarius* according to geographic location [14]. The black arrow points to sampling locality of Burgos within the distribution of *S. coronatus* mentioned in the text, and the white arrow indicates Illano at the contact of *S. coronatus* and *S. granarius* (see Supplementary Table S2).
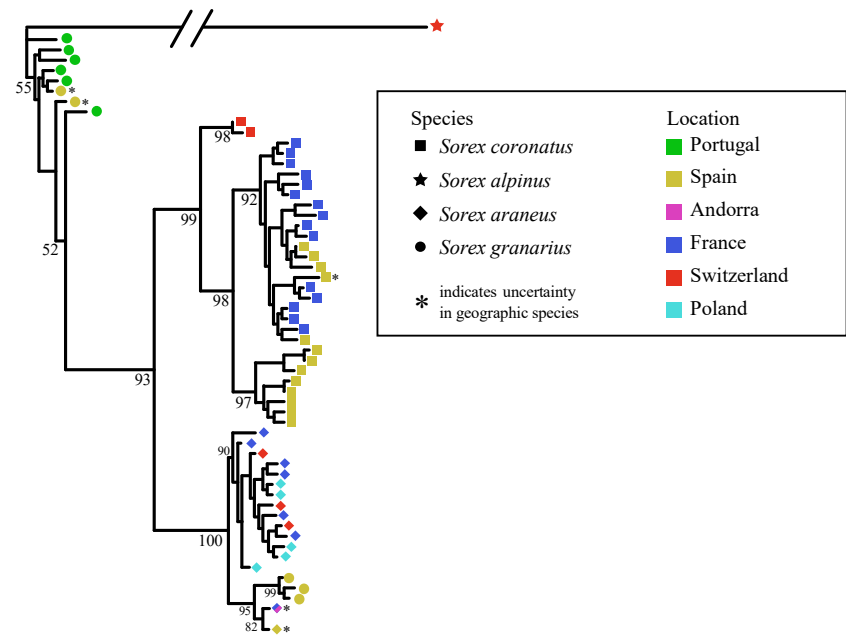
**Figure 2.** Maximum likelihood phylogenetic tree for *CytB* sequences of members of the *Sorex araneus* complex from southwestern Europe. Only haplotypes are shown, with geographic location and species indicated by symbols. *S. alpinus* is the outgroup. Six individuals were from geographic regions where two species occur; asterisks indicate their haplotypes. Their species designation is justified in the text. Bootstrap support over 1000 replicates is indicated on major internal nodes.
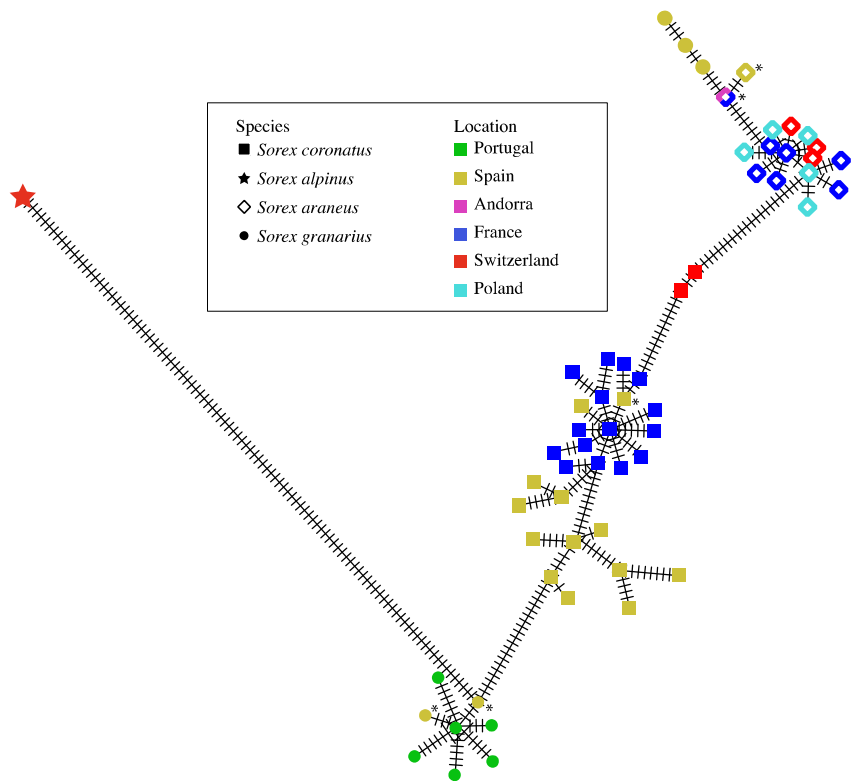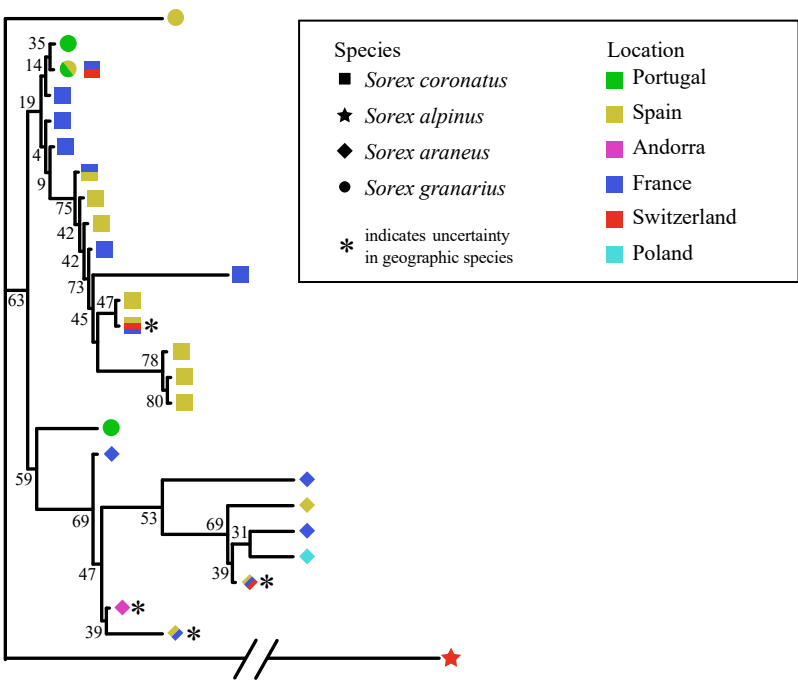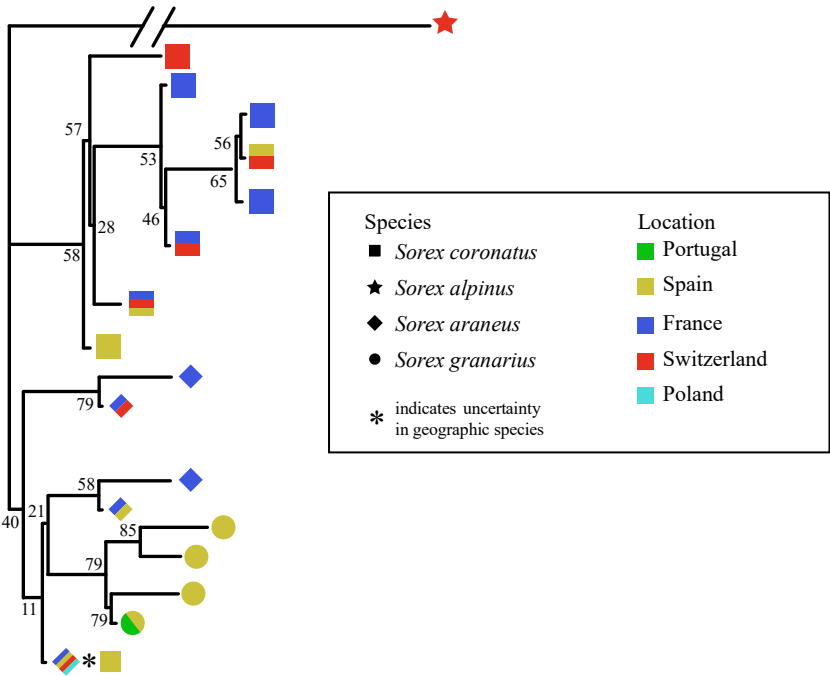


**Figure 3.** Minimum spanning network for *CytB* sequences of members of the *Sorex araneus* complex from southwestern Europe. Only haplotypes are shown, with geographic location and species indicated by symbols. *S. alpinus* is the outgroup. Bars on branches indicate mutational steps. Six individuals were from geographic regions where two species occur; asterisks indicate their haplotypes. Their species designation is justified in the text.

**Figure 4.** Maximum likelihood phylogenetic tree for *ApoB* sequences of members of the *Sorex araneus* complex from southwestern Europe. Only haplotypes are shown, with geographic location and species indicated by symbols. *S. alpinus* is the outgroup. Six individuals were from geographic regions where two species occur; asterisks indicate their haplotypes. Their species designation is justified in the text. Bootstrap support over 1000 replicates is indicated on major internal nodes.



**Figure 5.** Maximum likelihood phylogenetic tree for *BRCA1* sequences of members of the *Sorex araneus* complex from southwestern Europe. Only haplotypes are shown, with geographic location and species indicated by symbols. *S. alpinus* is the outgroup. An asterisk indicates a haplotype found in an individual (*S. araneus*) from a geographic region where *S. araneus* and *S. coronatus* occur; the same haplotype also occurs in individuals away from the contact area whose other markers indicate that they are *S. coronatus*. Bootstrap support over 1000 replicates is indicated on major internal nodes.
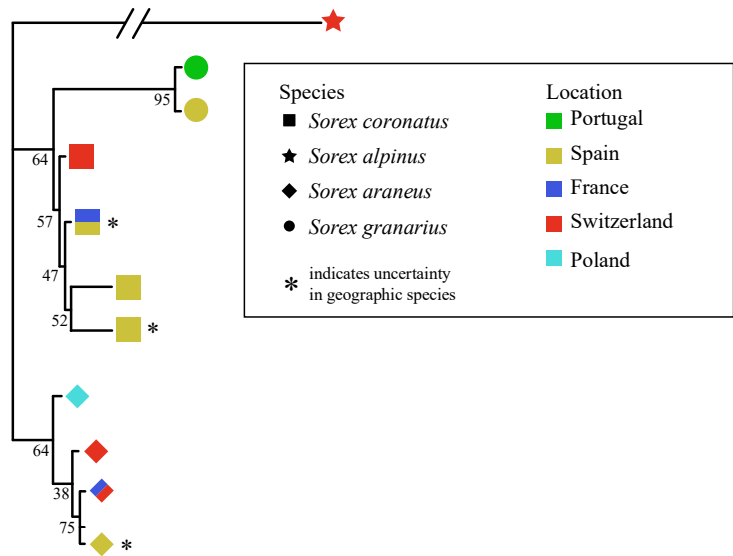
**Figure 6.** Maximum likelihood phylogenetic tree for *DBY7* sequences of members of the *Sorex araneus* complex from southwestern Europe. Only haplotypes are shown, with geographic location and species indicated by symbols. *S. alpinus* is the outgroup. Three individuals were from geographic regions where two species occur; asterisks indicate their haplotypes. Their species designation is justified in the text. Bootstrap support over 1000 replicates is indicated on major internal nodes.
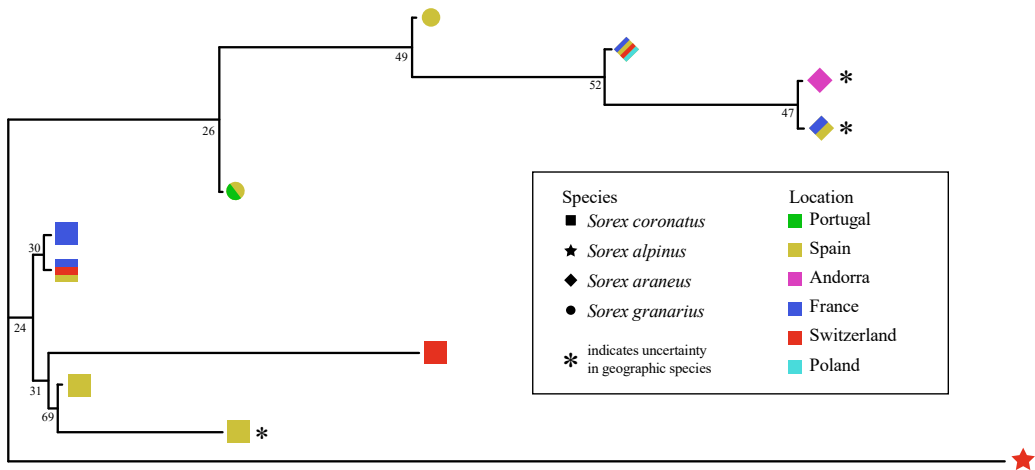


**Figure 7.** Maximum likelihood phylogenetic tree for *ZFX1* sequences of members of the *Sorex araneus* complex from southwestern Europe. Only haplotypes are shown, with geographic location and species indicated by symbols. *S. alpinus* is the outgroup. Four individuals were from geographic regions where two species occur; asterisks indicate their haplotypes. Their species designation is justified in the text. Bootstrap support over 1000 replicates is indicated on major internal nodes.

## 4. Discussion

Yannic et al. [13,14] inferred that the *S. granarius* from the Central System mountain range of Spain had a mitochondrial genome derived from *S. araneus*, specifically from a population currently represented by *S. araneus* in the Pyrenees (the Carlit chromosomal race). They made this inference of mitonuclear discordance based on the very close similarity of the *S. granarius CytB* sequence to that of the Pyrenean *S. araneus*. In contrast, a range of nuclear markers showed *S. granarius* and *S. araneus* to be well-distinct phylogenetically. Thus, they presumed that *S. granarius* ancestrally had a distinctive nuclear and mitochondrial genome and that it was only on hybridization with *S. araneus* that they lost their own mitochondrial genome and gained that of *S. araneus*. However, the existence of a distinctive "original" *S. granarius* mitochondrial genome was only conjecture – they did not demonstrate it.

Here we show that there is a distinctive mitochondrial genome that can be ascribed to *S. granarius* from the large continuous area of Portugal and northwestern Spain. This is demonstrated by *CytB* sequences of eight individuals from seven localities widely spaced in this area of western Iberia, within the reported distribution area of *S. granarius* (Supplementary Table S2; Figure 1). Four of these individuals from four localities have haplotypes at 2-4 nuclear loci that are phylogenetically close to the *S. granarius* from the Central System. This suggests that *S. granarius* has both a distinctive nuclear and mitochondrial genome that is found in these four individuals. The other four individuals from western Iberia with *CytB* sequences of the *S. granarius* type did not have any nuclear data.

Thus, our results confirm Yannic et al.'s [13,14] inference that *S. granarius* "captured" the mitochondrial genome of *S. araneus*, with the original mitochondrial genome of *S. granarius* (that we described in western Iberia) being replaced by that of *S. araneus* in those nine individuals from three localities in the Central System mountain range of Spain described by Yannic et al. [14].

How might this mitochondrial capture have occurred? It must have involved hybridization between *S. granarius* and *S. araneus* when one or both species had a wider distribution and were in contact, since currently they are allopatric (Figure 1). Thus, the *S. granarius* with the captured mitochondrial genome are found in the Central System of Spain far distant from the current population of *S. araneus* with the closely similar mitochondrial genome in the Pyrenees. It seems unlikely that the introgression event occurred during the Holocene. In a European context, all *Sorex* species are found in places with relatively cool, damp conditions [15]. Thus, in southern Europe they occur largely in mountainous areas; the exception being the coast around northern and northwestern Iberia occupied by *S. coronatus* and *S. granarius* (Figure 1). In an Iberian context, this is a relatively cool, wet area. *Sorex* do not occur in the hot dry conditions of central and southern lowland Iberia. Therefore, it can be presumed that the *S. granarius* in the Central System of Spain have long been separated from the *S. araneus* in the Pyrenees. It appears most likely that the two species were in contact during the Last Glaciation when there were cooler damper conditions in Iberia [24]. Given its current restriction there, it can be suggested that *S. granarius* was localized in Iberia throughout the Last Glaciation. *S. araneus* is believed to have had a glacial refugium in southeastern Europe and spread widely eastwards, westwards and northwards in warmer conditions after the Last Glacial Maximum (LGM)[25] and presumably got to southwestern Europe at that time. It is difficult to be exactly sure of the ranges of *S. araneus*, *S. coronatus* and *S. granarius* during the Late Glacial period from the end of the LGM until the beginning of the Holocene (comprising the time interval of approximately 20 – 10,000 years ago, incorporating the Bølling–Allerød warm period followed by the Younger Dryas cold period [24]). Both climatic conditions and their competitive interactions, including competitive exclusion, would have been important [26,27]. *S. araneus* and *S. granarius* could easily have had different ranges than at present in Iberia and come into contact, creating the opportunity for hybridization and introgression. The current distributions of the three species, with *S. araneus* and *S. granarius* not only disjunct but with the Carlit race of *S. araneus* in the Pyrenees completely surrounded by *S. coronatus* (Figure 1), indicates a dynamism of range changes during the Bølling–Allerød, Younger Dryas and early Holocene. Other examples of strange disjunct distributions deriving from range changes at this time can be seen in phylogeographic studies of other species, e.g. the distribution of the Carpathian and Eastern mitochondrial clades in the bank vole *Clethrionomys glareolus* in central, northern and western Europe [28]. Clues as to how *S. araneus* and *S. granarius* could have come into contact comes from their current distributions and that of *S. coronatus* (Figure 1). The relative climatic tolerance of *S. araneus* and *S. coronatus* is particularly clear. Thus, where the two species are in parapatry in southwestern Europe, *S. araneus* occurs in the cooler, damper, higher altitude areas (i.e. the Pyrenees, Massif Central). The greater tolerance of cooler, wetter conditions by *S. araneus* over *S. coronatus*, and vice versa for warmer, drier conditions, is even evident at a microgeographic level, based on studies in Switzerland [26]. Thus, through competitive exclusion, it can be conjectured that once *S. araneus* had spread to southwestern Europe after the

LGM, it actually may have had a wider range than *S. coronatus* in northern Iberia during the Younger Dryas cold phase. This may have been the time that it came into contact with *S. granarius* (allowing capture of the *S. araneus* mitochondrial genome by *S. granarius*) because *S. granarius* would likely have been able to occupy lowland central Iberia at that time (rather than being restricted to the mountains, as at present) because of the overall cooler conditions. Hence, *S. araneus* expanding southwards and *S. granarius* northwards could have met. At the end of the Younger Dryas, the warmer conditions would presumably have caused extinction of *Sorex* from lowland central Iberia, with *S. granarius* (with a *S. araneus* mitochondrial genome) retreating to the Central System mountain range (Figure 1). The warmer conditions of the early Holocene apparently favoured the spread of *S. coronatus* at the cost of *S. araneus* [29]. It came to occupy nearly the whole of northern Spain, France and nearby areas (Figure 1), spreading to northern France after 8000 years ago (because it did not get across the land bridge that connected to Britain until this time), but before 5000 years ago (because it did manage to spread to the Channel Isles which were connected to northern France until this time) [29]. It can be supposed that it was at the time of spread of *S. coronatus* that *S. araneus* became restricted to the high altitude populations in the Pyrenees and Massif Central. The *S. araneus* population in the Pyrenees may be considered the remnant population of that which came into contact with *S. granarius*.

There have been various suggestions as to how mitochondrial capture could be a neutral process and merely reflect differing population dynamics of the two interacting populations [e.g. 30]. While such a neutral process may explain the mitochondrial capture of the *S. araneus* mitochondrial genome by *S. granarius* in Iberia, there is actually no a priori expectation of vastly different population size or spread between *S. araneus* and *S. granarius*. However, it is striking that in a number of mammals in Europe – *Clethrionomys* in Fennoscandia and the Urals [28] and *Lepus* in Iberia [31], relatively warm-tolerant species have captured the mitochondrial genomes of relatively cold-tolerant species. *Sorex* in Iberia follows a similar pattern and, together with the other examples of mitochondrial introgression from a cold-tolerant to a warm-tolerant species may indicate that the captured mitochondrial genome provides selectively favourable traits [5]. Thus, the mitochondrial genome is related to energy production, so it may be expected that a cold-adapted species would have a cold-adapted mitochondrial genome. Clearly further bioinformatics studies of the whole mitochondrial genome and follow-up physiological studies would be necessary to confirm this suggestion. It is also worth noting that an additional case of mitochondrial capture has been described for Iberia involving wall lizards [32], so it would also be worth thinking whether there is something specific about the conditions in Iberia that make it more likely that mitochondrial capture happens there.

That *S. granarius* and *S. araneus* were able to hybridize is striking because every species within the *S. araneus* complex has a different karyotype, and karyotypic difference is believed important in promoting reproductive isolation in the group [33]. The Carlit chromosomal race of *S. araneus* has one of the simplest, least derived karyotypes of any chromosomal race in the species [11], and this may increase the chance of successful hybridization. The Carlit chromosomal race of *S. araneus* ($2n = 26$-$29$) differs from *S. granarius* ($2n = 36$-$37$) due to occurrence of Robertsonian (Rb) fusions - with four-five metacentric chromosomes found in *S. araneus* which are present as acrocentrics in *S. granarius* [11]. In hybrids, the Rb metacentrics from *S. araneus* will, in each case, pair and recombine with, and segregate from, two homologous acrocentrics from *S. granarius*, forming trivalent configurations at meiosis I [34]. For a hybrid to have four-five such meiotic trivalents would not be expected to promote particularly high levels of infertility, on chromosomal grounds alone [34,35]. However, given that *S. granarius* and *S. araneus* have been geographically isolated from each other and accumulated genic differences, hybrids between the species may suffer an unfitness deriving from both genic and chromosomal factors. There is a precedence in the *Sorex* shrews in Switzerland where naturally occurring hybrids between *Sorex antinorii* and *S. araneus* have an unfitness considerably greater than expected from chromosomal difference alone [36]. However, in that case, the chromosomal difference is

substantially more than between the Carlit chromosomal race of *S. araneus* and *S. granarius*, and even then reproductive isolation is not absolute, allowing a degree of introgression [37]. Therefore, when *S. araneus* and *S. granarius* came into contact, it is reasonable to expect that they would have been able to mate with each other, and that the hybrids, even if unfit, may have produced occasional offspring, thus permitting genetic introgression, including mitochondrial capture.

What about the current contacts between *Sorex* in Iberia? Based on the published distribution map (reproduced in Figure 1), *S. coronatus* and *S. araneus* are currently sympatric or parapatric in the Pyrenees, while *S. coronatus* and *S. granarius* are geographically close and possibly in contact in northwestern Spain. On chromosomal grounds, knowing that there can be some successful hybridization and introgression between *S. araneus* and *S. granarius* and between *S. antinorii* and *S. araneus*, then it is a possibility that there may be hybridization and introgression between *S. coronatus* and *S. araneus* and between *S. coronatus* and *S. granarius*. This is particularly the case for *S. coronatus* ($2n = 22\text{-}23$) and *S. granarius* ($2n = 36\text{-}37$), with seven metacentrics found in *S. coronatus* present as acrocentrics in *S. granarius* [11], together with three additional non-Rb chromosomal rearrangements distinguishing the two species [38]. This is a greater difference than between *S. granarius* and the Carlit chromosomal race of *S. araneus*, but not massively so. For *S. coronatus* and the Carlit chromosomal race of *S. araneus*, the chromosomal difference is very substantial [11,38].

Indeed, our genetic data do suggest that there may have been hybridization and introgression between all pairs of *Sorex* species in Iberia. Thus, in Illano in northwestern Spain (Figure 1) two of the four shrews only have data for *CytB* and are categorized as *S. granarius*, while two only have data for nuclear markers and are categorized as *S. coronatus* (Supplementary Table S2). If all four individuals represent the norm for this population, this could be an example of cytonuclear discordance, i.e. a *S. coronatus* nuclear background with a *S. granarius* mitochondrial genome. More data are needed to be certain of this. For *BRCA1*, it is striking that exactly the same haplotype as found in the Carlit chromosomal race of *S. araneus*, located in the Pyrenees, occurs in *S. coronatus* in northern Spain going west from the Pyrenees (Figure 5). Five out of 18 individuals from this region have the *S. araneus* haplotype for *BRCA1*, while the others have haplotypes phylogenetically within *S. coronatus*. All five individuals with the *S. araneus* haplotype for *BRCA1* are characterized as *S. coronatus* based on 3-4 additional genetic markers and are within the previously described distribution area of *S. coronatus* (Figure 1). It is difficult to suggest this as anything other than an introgression event, i.e. that *S. coronatus* and *S. araneus* hybridized and with subsequent backcrossing there are individuals with largely *S. coronatus* genomes, but with a *S. araneus* allele at *BRCA1*. The fact that the *BRCA1* haplotype in *S. coronatus* is identical to the *BRCA1* haplotype in *S. araneus* suggests that the introgression event is recent (maybe during the late Holocene), and presumably occurred in the vicinity of the Pyrenees, with the *S. araneus* allele spreading westward thereafter. To fully understand the complexity of the introgression events involving *S. coronatus*, there is a need to sample the species more intensively across Spain – likewise with the neighbouring populations of *S. araneus* and *S. granarius*. Clearly the Iberian *Sorex* would be best studied using genomics, to be able to determine exactly which parts of the genome have introgressed. Such an approach has been adopted successfully in hares occurring in the Iberian Peninsula, and elsewhere [39].

## 5. Conclusions

Based on their previous studies, Yannic et al. [13,14] proposed a mitochondrial capture involving *Sorex* species in Iberia. Their inference was based on the substantial nuclear divergence and virtual identity in mitochondrial *CytB* sequence between *S. granarius* and *S. araneus*, with the hypothesis that *S. granarius* captured the *S. araneus* mitochondrial genome. Yannic et al. [13,14] had only genotyped *S. granarius* from a small part of their range. In our new sampling of *S. granarius*, we were able to find individuals which had both nuclear and mitochondrial genomes distinct from *S. araneus*. Thus, we believe that

we have found the ancestral mitochondrial genome of *S. granarius* – replaced by the *S. araneus* mitochondrial genome in populations from the Central System of Spain – providing the "missing piece" in the Yannic et al. [13,14] hypothesis. Furthermore, we provide a biogeographical scenario for the mitochondrial capture, and place the capture in the context of karyotypic differences between the species. We also discuss whether natural selection favoured *S. granarius* with the captured mitochondrial genome, but further studies are needed to address this more fully. Finally, our data suggest that there have been other introgression events among the three species of *Sorex* in Iberia.

**Supplementary Materials:** Table S1: Details of the loci, primers and annealing temperatures used in this study. Table S2: All *Sorex* from Portugal, Spain, France and Andorra that were genetically typed in this study for various combinations of the mitochondrial locus *CytB* and nuclear loci *ApoB*, *BRCA1*, *DBY7* and *ZFX1*. They were classified as *Sorex araneus*, *S. coronatus* or *S. granarius* based on geography [14] or genotype at nuclear or mitochondrial DNA loci (see text). Table S3: Diversity statistics for the mitochondrial locus *CytB* in *Sorex araneus*, *S. coronatus* or *S. granarius*, grouped according to nuclear or mitochondrial species identity (see Table S2). Data based on our new samples and those of Yannic et al. [13]. See text for further details.

# References

1. Moritz, C.; Dowling, T.E.; Brown, W.M. Evolution of animal mitochondrial DNA: relevance for population biology and systematics. *Ann. Rev. Ecol. Syst.* **1987**, *18*, 269–292.
2. Toews, D.P.; Brelsford, A. The biogeography of mitochondrial and nuclear discordance in animals. *Mol. Ecol.* **2012**, *21*, 3907–3930.
3. Tegelström, H. Transfer of mitochondrial DNA from the northern red-backed vole (*Clethrionomys rutilus*) to the bank vole (*C. glareolus*). *J. Mol. Evol.* **1987**, *24*, 218–227.
4. Ballard, J.W.O.; Whitlock, M.C. The incomplete natural history of mitochondria. *Mol. Ecol.* **2004**, *13*, 729–744.
5. Sloan, D.B.; Havird, J.C.; Sharbrough, J. The on-again, off-again relationship between mitochondrial genomes and species boundaries. *Mol. Ecol.* **2017**, *26*, 2212-2236.
6. Good, J.M.; Hird, S.; Reid, N.; Demboski, J.R.; Steppan, S.J.; Martin-Nims, T.R.; Sullivan, J. Ancient hybridization and mitochondrial capture between two species of chipmunks. *Mol. Ecol.* **2008**, *17*, 1313–1327.
7. Alves, P.C.; Melo-Ferreira, J.; Freitas, H.; Boursot, P. The ubiquitous mountain hare mitochondria: multiple introgressive hybridization in hares, genus *Lepus*. *Phil. Trans. R. Soc. B* **2008**, *363*, 2831–2839.
8. Melo-Ferreira, J.; Boursot, P.; Carneiro, M.; Esteves, P.J.; Farelo, L.; Alves, P.C. Recurrent introgression of mitochondrial DNA among hares (*Lepus* spp.) revealed by species-tree inference and coalescent simulations. *Syst. Biol.* **2012**, *61*, 367–381.

9.   Melo-Ferreira, J.; Seixas, F.A.; Cheng, E.; Mills, L.S.; Alves, P.C. The hidden history of the snowshoe hare, *Lepus americanus*: extensive mitochondrial DNA introgression inferred from multilocus genetic variation. *Mol. Ecol*. **2014**, *23*, 4617–4630.

10.  Seixas, F.A.; Boursot, P.; Melo-Ferreira, J. The genomic impact of historical hybridization with massive mitochondrial DNA introgression. *Genome Biol.* **2018**, *19*, 19.

11.  Bulatova, N.S.; Biltueva, L.S.; Pavlova, S.V.; Zhdanova, N.S.; Zima, J. Chromosomal differentiation in the common shrew and related species. In *Shrews, Chromosomes and Speciation*; Searle, J.B., Polly, P.D., Zima, J., Eds.; Cambridge University Press: Cambridge, UK, 2019; pp. 134–185.

12.  Wójcik, J.M.; Searle, J.B. The chromosome complement of *Sorex granarius*—the ancestral karyotype of the common shrew (*Sorex araneus*)? *Heredity.* **1988**, *61*, 225–229.

13.  Yannic, G.; Basset, P.; Hausser, J. A new perspective on the evolutionary history of Western European *Sorex araneus* group revealed by paternal and maternal molecular markers. *Mol. Phylogenet. Evol.* **2008**, *47*, 237–250.

14.  Yannic, G.; Dubey, S.; Hausser, J.; Basset, P. Additional data for nuclear DNA give new insights into the phylogenetic position of *Sorex granarius* within the *Sorex araneus* group. *Mol. Phylogenet. Evol.* **2010**, *57*, 1062–1071.

15.  Mitchell-Jones, A.J.; Amori, G.; Bogdanowicz, W.; Krystufek, B.; Reijnders, P.; Spitzenberger, F.; Stubbe, M.; Thissen, J.; Vohralik, V.; Zima, J. *Atlas of European Mammals*. Academic Press: London, UK, 1999.

16.  Thompson, J.D.; Higgins, D.G.; Gibson, T.J. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* **1994**, *22*, 4673–4680.

17.  Tamura, K.; Stecher, G.; Kumar, S. MEGA11: Molecular Evolutionary Genetics Analysis version 11. *Mol. Biol. Evol.* **2021**, *38*, 3022–3027.

18.  Stephens, M.; Smith, N.; Donnelly, P. A new statistical method for haplotype reconstruction from population data. *Am. J. Hum. Genet.* **2001**, *68*, 978-989.

19.  Rozas, J.; Ferrer-Mata, A.; Sanchez-DelBarrio, J.C.; Guirao-Rico, S.; Librado, P.; Ramos-Onsins, S.E.; Sanchez-Gracia, A. DnaSP 6: DNA sequence polymorphism analysis of large data sets. *Mol. Biol. Evol.* **2017**, *34*, 3299–3302.

20.  Nguyen, L.T.; Schmidt, H.A.; Von Haeseler, A.; Minh, B.Q. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol. Biol. Evol.* **2015**, *32*, 268–274.

21.  Paradis, E.; Schliep, K. ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. *Bioinformatics* **2019**, *35*, 526–528.

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

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

24.  Andersen, B.G.; Borns Jr, H.W. *The Ice Age World*; Scandinavian University Press: Oslo, Norway, 1997.

25.  Thaw, S.; White, T.A.; Bannikova, A.; Searle, J.B. Phylogeography. In *Shrews, Chromosomes and Speciation*; Searle, J.B., Polly, P.D., Zima, J., Eds.; Cambridge University Press: Cambridge, UK, 2019; pp. 112–133.

26.  Neet, C.R.; Hausser, J. Habitat selection in zones of parapatric contact between the common shrew *Sorex araneus* and Millet's shrew *Sorex coronatus*. *J. Anim. Ecol.* **1990**, *59*, 235–250.

27.  Neves, T.; Borda-de-Água, L.; Mathias, M.L.; Tapisso, J.T. The influence of the interaction between climate and competition on the distributional limits of European shrews. *Animals* **2021,** *12*, 57.

28.  Kotlík, P.; Marková, S.; Horníková, M.; Escalante, M.A.; Searle, J.B. The bank vole (*Clethrionomys glareolus*) as a model system for adaptive phylogeography in the European Theater. *Frontiers Ecol. Evol.* **2022**, *10*, 866605.

29.  Searle, J.B. Three new karyotypic races of the common shrew *Sorex araneus* (Mammalia: Insectivora) and a phylogeny. *Syst. Zool.* **1984**, *33*, 184–194.

30.  Currat, M.; Ruedi, M.; Petit, R.J.; Excoffier L. The hidden side of invasions: massive introgression by local genes. *Evolution* **2008**, *62*, 1908–1920.

31.  Melo-Ferreira, J.; Alves, P.C.; Freitas, H.; Ferrand, N.; Boursot, P. The genome legacy from the extinct *Lepus timidus* to the three hare species of Iberia: contrast between mtDNA, sex chromosomes and autosomes. *Mol. Ecol.* **2009**, *18*, 2643–2658.

32.  Renoult, J.P.; Geniez, P.; Bacquet, P.; Benoit, L.; Crochet, P.A. Morphology and nuclear markers reveal extensive mitochondrial introgressions in the Iberian Wall Lizard species complex. *Mol. Ecol.* **2009**, *18*, 4298–4315.

33.  Basset, P.; Yannic, G.; Hausser, J. Is it really the chromosomes? In *Shrews, Chromosomes and Speciation*; Searle, J.B., Polly, P.D., Zima, J., Eds.; Cambridge University Press: Cambridge, UK, 2019; pp. 365–383.

34.  Borodin, P.M.; Fedyk, S.; Chętnicki, W.; Torgasheva, A.A.; Pavlova, S.V.; Searle, J.B. Meiosis and fertility associated with chromosomal heterozygosity. In *Shrews, Chromosomes and Speciation*; Searle, J.B., Polly, P.D., Zima, J., Eds.; Cambridge University Press: Cambridge, UK, 2019; pp. 217–270.

35.  Searle, J.B. Chromosomal hybrid zones in eutherian mammals. In *Hybrid Zones and the Evolutionary Process*; Harrison, R.G., Ed.; Oxford University Press: New York, USA, 1993; pp. 309–353.

36.  Fedyk, S.; Pavlova, S.V.; Chętnicki, W.; Searle, J.B. Chromosomal hybrid zones. In *Shrews, Chromosomes and Speciation*; Searle, J.B., Polly, P.D., Zima, J., Eds.; Cambridge University Press: Cambridge, UK, 2019; pp. 271–312.

37.  Brünner, H.; Lugon-Moulin, N.; Hausser, J. Alps, genes, and chromosomes: their role in the formation of species in the *Sorex araneus* group (Mammalia, Insectivora), as inferred from two hybrid zones. *Cytogenet. Genome Res*. **2002**, *96*, 85–96.

38.  Biltueva, L.; Vorobieva, N.; Perelman, P.; Trifonov, V.; Volobouev, V.; Panov, V.; Ilyashenko, V.; Onischenko, S.; O'Brien, P.; Yang, F.; Ferguson-Smith, M. Karyotype evolution of Eulipotyphla (Insectivora): the genome homology of seven *Sorex* species revealed by comparative chromosome painting and banding data. *Cytogenet. Genome Res.* **2011**, *135*, 51–64.

39.  Ferreira, M.S.; Jones, M.R.; Callahan, C.M.; Farelo, L.; Tolesa, Z.; Suchentrunk, F.; Boursot, P.; Mills, L.S.; Alves, P.C.; Good, J.M.; Melo-Ferreira, J. The legacy of recurrent introgression during the radiation of hares. *Syst. Biol.* **2021**, *70*, 593-607.

40.  Irwin, D.M.; Kocher, T.D.; Wilson, A.C. Evolution of the cytochrome *b* gene of mammals. *J. Mol. Evol.* **1991**, *32*, 128–144.

41.  Jaarola, M.; Searle, J.B. Phylogeography of field voles (*Microtus agrestis*) in Eurasia inferred from mitochondrial DNA sequences. *Mol. Ecol.* **2002**, *11*, 2613–2621.

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

43.  Dubey, S.; Salamin, N.; Ohdachi, S.D.; Barrière, P.; Vogel, P. Molecular phylogenetics of shrews (Mammalia: Soricidae) reveal timing of transcontinental colonizations. *Mol. Phylogenet. Evol.* **2007**, *44*, 126–137.

44.  Dubey, S.; Zaitsev, M.; Cosson, J.F.; Abdukadier, A.; Vogel, P. Pliocene and Pleistocene diversification and multiple refugia in a Eurasian shrew (*Crocidura suaveolens* group). *Mol. Phylogenet. Evol.* **2006**, *38*, 635–647.