Molecular phylogenetics supports the origin of an endemic Balearic shrew lineage (*Nesiotites*) coincident with the Messinian Salinity Crisis

Pere Bover[a,b,c*], Kieren J. Mitchell[a], Bastien Llamas[a], Juan Rofes[d], Vicki A. Thomson[e], Gloria Cuenca-Bescós[f], Josep A. Alcover[b,c], Alan Cooper[a], Joan Pons[b]

[a] Australian Centre for Ancient DNA (ACAD), School of Biological Sciences, University of Adelaide, Australia
[b] Departament de Biodiversitat i Conservació, Institut Mediterrani d'Estudis Avançats (CSIC-UIB), Esporles, Illes Balears, Spain
[c] Research Associate, Department of Mammalogy/Division of Vertebrate Zoology, American Museum of Natural History, NY
[d] Archéozoologie, Archéobotanique: Sociétés, pratiques et environnements (UMR 7209), Sorbonne Universités, Muséum national d'Histoire naturelle, CNRS, CP56, 55 rue Buffon, 75005 Paris, France.
[e] School of Biological Sciences, University of Adelaide, Australia.
[f] Grupo Aragosaurus-IUCA, Universidad de Zaragoza, Spain.

* Corresponding author at: Australian Centre for Ancient DNA (ACAD), School of Biological Sciences, University of Adelaide, Darling Building, North Terrace Campus, Adelaide, SA, 5005, Australia (P. Bover).

E-mail addresses: pere.boverarbos@adelaide.edu.au (P. Bover), kieren.mitchell@adelaide.edu.au (K.J. Mitchell), bastien.llamas@adelaide.edu.au (B. Llamas), juan.rofes@mnhn.fr (J. Rofes), vicki.thomson@adelaide.edu.au (V. Thomson), cuencag@unizar.es (G. Cuenca-Bescós), jaalcover@imedea.uib-csic.es (J.A. Alcover), alan.cooper@adelaide.edu.au (A. Cooper), jpons@imedea.uib-csic.es (J. Pons).

Abstract

The red-toothed shrews (Soricinae) are the most widespread subfamily of shrews, distributed from northern South America to North America and Eurasia. Within this subfamily, the tribe Nectogalini includes the fossil species *Nesiotites hidalgo* recorded from the Late Pleistocene to Holocene of the Balearic Islands (Western Mediterranean). Although there is a consensus about the close relationship between the extinct red-toothed shrew genera *Nesiotites* and *Asoriculus* based on morphology, molecular data are necessary to further evaluate the phylogenetic relationships of the Balearic fossils. We obtained a near complete mitochondrial genome of *N. hidalgo*, allowing the first molecular phylogenetic analysis of this species. Analyses based on 15,167 bp of the mitochondrial genome placed *N. hidalgo* as close relative to the extant Himalayan shrew (*Soriculus nigrescens*), and a combined analysis using molecular and morphological data confirm that *N. hidalgo* and *Asoriculus gibberodon* are sister-taxa with *S. nigrescens* as the immediate outgroup. Molecular clock and divergence estimates suggest that the split between *N. hidalgo* and its closest living relative occurred around 6.44 Ma, which is in agreement with the previously proposed colonisation of the Balearic Islands from mainland Europe by nectogaline shrews during the Messinian Salinity Crisis (5.97–5.33 My ago). Our results highlight that it is possible to retrieve genetic data from extinct small mammals from marginal environments for DNA preservation. Additional finds from the fossil record of Soricinae from the Eurasian Late Miocene/Early Pliocene are needed to shed further light on the still confusing taxonomy and paleobiogeography of this clade.

Keywords: Ancient DNA, Fossil micromammal, Mallorca, Soricinae, Phylogeny

1. Introduction

Shrews (Soricidae) are small insectivores (< 100 grams) related to talpid moles and hedgehogs, and represent one of the most diverse families of mammals with 385 extant species in 26 genera (Wilson and Reeder, 1985). Living shrew species are widely distributed across Europe, Asia, Africa, and the Americas, and have adapted to a wide variety of habitats and lifestyles, including species that are fossorial or semi-aquatic (*e.g.*, Hutterer, 1985). The red-toothed shrews (Soricinae)—so named for the coloured iron compounds in their enamel—are the most widespread subfamily of shrews. However, the phylogenetic relationships among red-toothed shrew taxa are still poorly understood, making it challenging to reconstruct biogeographic or macroevolutionary patterns. This uncertainty also makes it difficult to assess the affinities of shrew species described from the fossil record, especially where the fossil species appear to have no close living relatives. A particularly contentious matter is the origin and evolution of several species of extinct red-toothed shrews that have been described from Mediterranean islands and initially assigned to the fossil genus *Nesiotites* (Bate, 1944): the Balearic shrew (*N. hidalgo*) from Mallorca and Menorca (Fig. 1A), the Corsican shrew (*N. corsicanus*), and the Sardinian shrew (*N. similis*). It has recently been demonstrated that ancient DNA can be isolated from subfossil remains of microfauna from environments marginal for DNA preservation similar to the Mediterranean (*e.g.*, Guimaraes et al., 2016), providing a new avenue for investigating this issue.

The fossil record of *Nesiotites* from Mallorca extends back to the Late Miocene (Bover et al., 2014), with a series of chronospecies being described: *N. hidalgo* (type species; Middle Pleistocene to Holocene) (Bate, 1944), an Early Pleistocene form [referred to as *N.* ex. interc. *ponsi-hidalgo* by Reumer (1981) or *N.* aff. *ponsi* by Alcover et al. (1981)], *N. ponsi* (Late Pliocene; Reumer, 1979), an Early-Late Pliocene form (*Nesiotites* sp.; Pons-Moyà, 1990), *N. rafelinensis* (earliest Early Pliocene; Rofes et al., 2012), and finally a soricine from the Late Miocene-Early Pliocene (Bover et al., 2014). The continuous history of *Nesiotites* in Mallorca suggests that its ancestors probably arrived during the Messinian Salinity Crisis (MSC, 5.97–5.33 My ago; Krijgsman et al., 1999; Manzi et al., 2013), when low sea levels in the Mediterranean connected the island to mainland Europe. Several other Balearic animal taxa also appear to have arrived during the MSC, including at least five other mammals, eight reptiles, and four amphibians (Bover et al., 2014; Torres-Roig et al., 2017). Molecular clock studies have supported the colonisation of the Balearic Islands during the MSC by toads (*e.g.*, Arntzen et al., 1995; Fromhage et al., 2004) and some invertebrate species (*e.g.*, Chueca et al., 2017; Mora et al., 2017). However, the three endemic mammal lineages that survived until the Holocene—*Nesiotites*, *Myotragus* (Bovidae), and *Hypnomys*

(Gliridae)—became extinct after the arrival of the first human settlers around 4,300 years ago (Bover and Alcover, 2003, 2008; Bover et al., 2016). Marginal conditions for ancient DNA preservation in subfossils from the Mediterranean means that the potential MSC associated origins of these lineages remains to be tested with molecular data.

The exact phylogenetic and biogeographic origins of *Nesiotites* are also debated. Bate (1944) suggested a phylogenetic affinity of *Nesiotites* with the semi-aquatic Asian water-shrews belonging to the genera *Chimarrogale* and *Nectogale*. Conversely, Reumer (1980, 1984) proposed a close relationship between *Nesiotites* and terrestrial shrews from the genus *Episoriculus*. While all extant *Episoriculus* species are distributed in Asia, some putative fossil species—since reassigned to the fossil genus *Asoriculus* (Hutterer, 1994)—have been described from Europe. However, Reumer (1998) considered *Episoriculus* a subgenus of *Soriculus*, differentiating it from *Asoriculus*. A consensus has emerged based on analyses of morphological data that *Nesiotites* and *Asoriculus* are closely related (*e.g.*, Alcover et al., 1981; Masini and Sarà, 1998; Maul and Rzebik-Kowalska, 1998; Pons-Monjo et al., 2010, 2012; Rofes and Cuenca-Bescós, 2006, 2009), with some authors considering *Nesiotites* an insular descendent of the fossil taxon *Asoriculus gibberodon* (*e.g.*, Rofes and Cuenca-Bescós, 2009). However, under that framework the genus *Nesiotites* as applied by Bate (1944) is polyphyletic, so herein we use *Nesiotites* to refer exclusively to the Balearic clade (*i.e.*, excluding the extinct Corsican and Sardinian shrews). However, since the affinities of *Asoriculus* are debated [see Rofes et al. (2012) for a synthesis], this does not clarify the phylogenetic position of *Nesiotites* more broadly. Molecular data will be necessary to disentangle this issue.

We sequenced a near complete mitochondrial genome from a specimen of *N. hidalgo* and compared these data to previously published sequences from extant shrews. We augmented this molecular dataset with a morphological character matrix comprising new and existing data from a number of living and extinct shrew species (including both *Nesiotites* and *Asoriculus*). Using these data, we tested the temporal, phylogenetic, and biogeographic origins of *Nestiotites*, and refined the phylogeny of the red-toothed shrews.

2. Materials and Methods

2.1. Samples

Seven different *Nesiotites hidalgo* mandibles were collected in 2012 during exploration or survey campaigns of two Mallorcan caves (Fig. 1). Six mandibles came from a deposit in Cova des Garrover (Alcúdia), a cave in the north of Mallorca from which the most recent radiocarbon dated remains of *N. hidalgo* have been recovered

(Beta 163133, 4280±50, 3030–2690 cal BC; Bover and Alcover, 2008). In Cova des Garrover, *N. hidalgo* remains are located on the surface layer mixed with human-introduced vertebrate species (*e.g.*, the wood mouse *Apodemus sylvaticus*). The mass of most samples was low, so in order to reach a minimum amount of bone for DNA extraction we made two pools of three mandibles each: samples ACAD13280, ACAD13281, and ACAD13285 (three left mandibles, hereafter pool A); and ACAD13286, ACAD13289, and ACAD13292 (one left and two right mandibles, hereafter pool B) (Fig. 1B and 1C).

The third extracted sample is a single right mandible from Coveta des Gorgs, ACAD14872 (Escorca, Mallorca, Fig. 1A and 1D). This specimen was obtained from the surface level of the cave together with fossil *Myotragus balearicus* (Bovidae) remains. Radiocarbon dated *M. balearicus* bones from Coveta des Gorgs indicate a chronology range from 4456±33 BP (RICH-21771, 3340–3010 calBC) to 8650±40 (Beta-1431 17, 7651–7534 calBC; Bover and Alcover, 2003; Bover et al., 2016; Lalueza-Fox et al., 2002) (Table S1). The mandible of *N. hidalgo* was entirely consumed during DNA extraction and it could not therefore be directly radiocarbon dated. The wide range of dates obtained for *M. balearicus* bones coming from the same surface level of this cave means we cannot be certain of the precise age of our *N. hidalgo* sample, and it should be considered approximately 5,000 to 9,500 years old.

All pre-treatment, extraction and library preparation steps were performed at the dedicated ancient DNA laboratory of the Australian Centre for Ancient DNA (ACAD), University of Adelaide, Australia. The small size of the samples precluded the removal of a surface layer of the bone using mechanical tools, so the preparation of samples was performed following two alternative methods. Surface dirt was removed from Pool A specimens and ACAD14872 using sterile surgical blades, and the bones were then irradiated with UV for 30 minutes (min) on each side, wiped with 3% sodium hypochlorite, soaked for 2 min in 80% ethanol to fully remove bleach, air-dried, and finally UV irradiated again for 15 min on each side. To evaluate the possible effect of the sodium hypochlorite cleaning on the preservation of the DNA in the bone, mandibles of Pool B were just cleaned using surgical blades and irradiated with UV for 15 min on each side.

2.2. DNA extraction, library preparation, and hybridization enrichment

Bone specimens were powdered using a Braun Mikrodismembrator U (B. Braun Biotech International, Germany) by placing them inside a sterilised stainless steel container with an 8 mm tungsten ball for 5 seconds (s) at 3000 revolutions per minute (rpm). The amount of bone powder obtained for each sample is as follows: 210 mg for

Pool A, 180 mg for Pool B, and 60 mg for ACAD14872. Bone powder of Pools A and B were both decalcified and digested overnight at 55 ºC on a rotary wheel in 4.24 mL of decalcification/digestion buffer consisting of 0.47 M EDTA (pH 8.0), 0.47 % SDS, and 0.18 mg/mL Proteinase K. The same final reagent concentrations were used for ACAD14872, but in half the volume. DNA binding was performed using modified QG buffer (Qiagen) and a suspension of silicon dioxide particles (see Brotherton et al. 2013). Samples were then purified using 80 % ethanol and eluted in TLE (10 mM Tris, 0.1 mM EDTA ph 8.0) obtaining a final volume of 200 µL of extract (100 µL in the case of ACAD14872). A negative control was extracted along with each batch of samples.

PCR screening for the mtDNA 12S gene was used to test for the presence of DNA in the samples and extraction blank controls. Primers Mamm12S_E (Forward) and Mammal12S_H (Reverse) (Macqueen et al. 2010) were used to amplify a 96–97 bp 12S fragment. Two microlitres of template were used in each reaction in a final volume of 25 µL containing: 1 × Platinum Taq High Fidelity Buffer (Invitrogen), 3 mM $MgSO_4$, 0.4 µM each primer, 0.25 mM each dNTP, 1.25 U Platinum Taq HiFi (Invitrogen), 2 mg/mL Rabbit Serum Albumin (RSA, Sigma-Aldrich), and sterile water. PCR cycling conditions were: initial denaturation at 94 ºC for 2 min; 50 cycles of denaturation at 94 ºC for 20 s; primer annealing at 55 ºC for 15 s; elongation at 68 ºC for 30 s; and a final elongation step at 68 ºC for 10 min. PCR products were separated using gel electrophoresis on a 3.5 % agarose gel, stained with Gel-Red (Jomar Bioscience), and subsequently visualised under UV light. PCR products were purified using Agencourt AMPure XP magnetic beads (Beckman Coulter) according to the manufacturers protocol. Both the forward and reverse strands were sequenced using the BigDye Terminator Kit (Applied Biosystems); dye terminator was removed using Agencourt CleanSEQ magnetic particle solution (Beckman Coulter) and fragments were examined on a 3130xl Genetic Analyzer (Applied Biosystems).

Positive PCR amplification of a fragment 97 bp long was obtained from Pool B and ACAD14872. The best match for the nucleotide sequence obtained from Pool B using the BLASTn algorithm (Altschul et al. 1990) was sheep (*Ovis aries*, coverage 100 %, identity 100 %), whereas the best match for the sequence obtained from ACAD14872 was the Taiwanese brown-toothed shrew (*Episoriculus fumidus*, coverage 100 %, identity 94 %). The presence of sheep and goats dwelling inside Cova des Garrover could explain the amplification of sheep sequences in Pool B, with the sodium hypochlorite treatment of Pool A possibly removing this contamination and explaining our negative PCR result. It is also possible that the sheep DNA represents low-

concentration reagent contamination as mammalian DNA (particularly domestic species are known contaminants in PCR reagents (Leonard et al., 2007).

Following the successful PCR amplification of shrew DNA from ACAD14872, we constructed a double-stranded DNA library from this sample using a protocol based on Meyer and Kircher (2010) with several modifications outlined by Llamas et al. (2016). In brief, these modifications comprised the inclusion of a 5-mer barcode sequence in the truncated P5 adapter, the use of Platinum Taq Hifi (Invitrogen) for the first post-Bst PCR library amplification, and the use of Amplitaq Gold (Life Technologies) for subsequent amplifications. We enriched this library for mammal mitochondrial DNA using hybridisation capture with commercially synthesised RNA probes (Mycroarray, http://www.mycroarray.com/mybaits) following the protocol of Mitchell et al. (2016). Following enrichment, we added full-length Illumina sequencing adapters with a final round of PCR and the library was sequenced on an Illumina HiSeq2500 Rapid Run (2 x 100 bp paired-end).

2.3. Sequencing data filtering, mapping, and draft mitochondrial genome

The quality of resulting sequencing reads was analysed using FastQC v0.11.2 (http://www.bioinformatics.babraham.ac.uk/projects/fastqc). Using Sabre v.1.0 (http://github.com/najoshi/sabre) we filtered for samples containing the correct P5 5-mer barcode sequence (allowing one mismatch). Adapter sequences were trimmed with AdapterRemoval v.2.1.7 (Schubert et al., 2016) using the following parameter values: mismatch rate 0.1, minimum Phred quality 4, quality base 33, trim ambiguous bases (N), and trim bases with qualities equal or smaller than the given minimum quality. Paired reads overlapping by at least 11 bp were collapsed (merged) into a single read using AdapterRemoval. After trimming, reads shorter than 25 bp were excluded from further analysis.

We constructed a draft mitochondrial genome sequence for ACAD14872 using the MITObim pipeline v.1.8 (Hahn et al., 2013). This pipeline maps all reads to a reference seed using MIRA (Chevreux et al., 1999) and then iteratively attempts to extend mapped fragments. Using default parameters in MITObim, we mapped all 1,768,862 collapsed reads to the published mitochondrial genome sequence for *Episoriculus fumidus* (NC_003040), which was the species with the best hit for the short Sanger sequenced 12S sequence in the BLASTn analysis. Following 10 iterations of mapping, we converted the MAF formatted output to SAM file format using ‘miraconvert’ in MIRA, removed duplicate reads using ‘FilterUniqueSAMCons.py’ (Kircher, 2012), and generated a majority-rule consensus sequence from the retained reads using Geneious v9.1.3 (Biomatters, http://www.geneious.com, Kearse et al.,

2012). We then re-mapped all collapsed reads against this draft consensus using BWA v.0.7.13 (Li and Durbin, 2009) with established ancient DNA parameter values: -l 1024 (seed inactivated), -n 0.01, -o 2. Reads with mapping quality lower than a Phred score 25 were removed using SAMtools v.1.3.1 (Li et al., 2009) and duplicate reads were filtered using 'FilterUniqueSAMCons.py'. A final 75% majority consensus sequence was then generated in Geneious using 21,841 retained reads, calling nucleotides only at sites with read-depth ≥ 3x. The final consensus coverage was 95.6% of the reference (minus a 1,000 bp fragment of the control region) with a mean coverage depth of 124x (Accession Number MH010393).

In order to verify our MITObim consensus sequence, we also attempted to build a consensus using an iterative BWA mapping approach. Collapsed reads were mapped to the *Episoriculus fumidus* (NC_003040) mitochondrial genome using the BWA options mentioned above, and a 75% majority consensus was generated using Geneious where the reference sequence was used for sites with coverage depth < 3x. The newly generated consensus was then used as reference for a subsequent round of mapping, and this process was iterated until no additional reads were mapped. After 27 iterations, we were able to map 20,926 unique *Nesiotites* reads (86.3% of reference covered, mean coverage depth of 119x), and a final consensus sequence was generated as above (75% majority, calling nucleotides with read-depth ≥ 3x). Our MITObim and BWA consensus sequences were identical in the regions where they could be compared (*i.e.*, 15,073 nucleotides), but our BWA consensus displayed two additional gaps in the coding region spanning 885 nucleotides, which were successfully filled by our MITObim mapping. Thus, we conclude that our final *Nesiotites* consensus is robust to the mapping strategy chosen.

Nucleotide misincorporation and DNA fragmentation patterns in mapped reads were assessed using mapDamage v2.0.2 (Jónsson et al., 2013). MapDamage analysis (Fig. S1) displays the expected damage pattern observed in ancient samples (*e.g.*, Briggs et al., 2007). The truncated pattern of the curve observed at the 3' extreme is the result of trimming of the first nucleotide of all reverse reads (due to the presence of an A in the first position of all reverse reads).

As expected for mammals, the mitochondrial genome of *N. hidalgo* includes 13 protein-coding genes, two rRNA genes, 22 tRNA genes, and one control region. Most protein-coding genes start with the codon ATG, with the exception of *ND2* and *ND5* (ATA) and *ND3* (ATT). Termination codons for the protein-coding genes display the same pattern as other nectogaline shrews (*e.g.*, Fu et al., 2016; Huang et al., 2014; Liu et al., 2015), *i.e.*, TAA for *CO1*, *CO2*, *ATP8*, *ATP6*, *ND3*, *ND4L*, *ND5* and *ND6*; TAG for

ND1 and *ND2*; a truncated stop codon (T-- or TA-) for *CO3* and *ND4*; and AGA for *CYB.*

2.4. Phylogenetic analyses

Using Geneious, we aligned our *N. hidalgo* mitochondrial genome sequence with 15 previously published mitochondrial genome sequences from extant shrews and individual mitochondrial gene sequences for eight additional species (Table S2). The total alignment length was 15,167 bp. Stems and loops of mitochondrial rRNAs and tRNAs were identified using RNA-alifold (Bernhart et al., 2008), and ambiguously aligned regions of 12S rRNA and 16S rRNA were removed from further analysis. Optimal partitioning schemes and substitution models (Table S3) were estimated in PartitionFinder v1.1.1 (Lanfear et al., 2012) for downstream phylogenetic analysis in MrBayes (Ronquist et al., 2012) and BEAST (Drummond et al., 2012).

Our partitioned MrBayes v3.2.3 (Ronquist et al., 2012) analysis employed four separate runs of four Markov chains each (one cold and three incrementally heated) using default priors. Each chain ran for 10 million generations sampling every 10,000 generations. Convergence in topology was assessed using the average standard deviation of clade (split) frequencies (< 0.02), while convergence in individual parameter values was assessed through broadly overlapping distributions and effective sample sizes > 200 in Tracer v1.6 (http://tree.bio.ed.ac.uk/software/tracer/). Sampled trees were summarized as a majority-rule consensus tree after discarding the first 10% of trees as burnin.

We also ran a combined analysis of molecular and morphological data in MrBayes, by adding a morphological matrix as an eleventh partition. This morphological matrix comprised 48 morphological characters for six species (including *Nesiotites* and *Asoriculus*) published by Rofes and Cuenca-Bescós (2009) plus 14 additional individuals from four species newly scored for this study [Vertebrate Collections of the Muséum National d'Histoire Naturelle of Paris (France): *Anurosorex squamipes* (CG 1997-90 and CG 1870-567), *Soriculus nigrescens* (CG 1981-1283 and CG 1981-1285 to 1288), *Nectogale elegans* (CG 1900-596 to 598), and *Chimarrogale himalayica* (CG 1900-594, CG 1900-595, CG 1895-533 and CG 1895-534), Table S4]. Our morphological partition was analysed using the Mkv model (Lewis 2001) with the correction for ascertainment bias (coding=informative) and equal rates across sites. Chains for this combined analysis were run for 40 million generations sampling every 40,000 generations, but the analysis was otherwise identical to that of the molecular data alone.

We performed a partitioned molecular dating analysis using BEAST v.1.8.4, with a Birth-Death tree prior and a single lognormal relaxed clock (with a rate multiplier parameter to allow individual partitions to evolve at different relative rates). Following the available fossil information (Reumer, 1989, 1994), we calibrated our phylogeny by constraining the age of the tree root in accordance with the presence of ancestral Soricinae/Crocidurinae in the Miocene (*e.g.*, Dubey et al., 2007; Fumagalli et al., 1999; He et al., 2010; Quérouil et al., 2001) using a lognormal distribution such that 95% of the prior distribution fell between 20.15–26.83 Ma with a median of 21 Ma (mean="0.0", stdev="0.98", offset="20.0"). Four analyses were run for 100 million generations sampling every 10,000 generations, and the first 10 % of samples were discarded as burnin before summarising the results. Convergence was assessed through broadly overlapping distributions between chains and effective sample sizes > 200 as calculated using Tracer v1.6. Output from individual runs were combined using LogCombiner v.1.8.4 and a final maximum clade credibility tree was generated using TreeAnnotator v.1.8.4 (Rambaut et al., 2010).

3. Results

The MrBayes analysis of our molecular dataset (Fig. S2) yielded a tree that was generally consistent with previously published phylogenies based exclusively on mitochondrial data (*e.g.*, He et al., 2010; Ohdachi et al., 2006; Willows-Munro and Matthee, 2011). Our results recapitulate the monophyly of the soricine tribes Blarinellini, Soricini, Anourosoricini and Nectogalini, and suggest a close relationship between Anourosoricini and Nectogalini. Within Nectogalini, we find well-supported clades, including a clade comprising the genus *Chodsigoa*, a clade comprising *Chodsigoa* and the sampled *Episoriculus* species (with the exception of *E. fumidus*), a clade comprising the water shrews *Chimarrogale himalayica* and *Nectogale elegans*, and a clade comprising *Soriculus nigrescens* and *Nesiotites hidalgo*. Our analyses show weak support for a dichotomy between all sampled water shrews (*Chimarrogale*, *Nectogale*, and *Neomys*) and a clade comprising *Soriculus*, *Nesiotites*, *Chodsigoa* and the sampled *Episoriculus* species (with the exception of *E. fumidus*), with *E. fumidis* forming the immediate outgroup. The position of *E. fumidus* as basal to Nectogalini suggests the genus *Episoriculus* is polyphyletic, as already observed by other authors (He et al., 2010; Ohdachi et al., 2006; Willows-Munro and Matthee, 2011).

The addition of morphological data to our MrBayes analysis did not greatly alter the topology and support values compared to the molecular analysis alone (Fig. 2). However, we did recapitulate a monophyletic *Neomys* (PP=0.96), and confirm that *N.*

and *A. gibberodon* (PP=0.93) are sister-taxa, with *S. nigrescens* as their immediate outgroup (PP=0.93).

Our BEAST analysis (Fig. 3) provided greater support for several nodes compared to our MrBayes analysis [*Chimarrogale+Nectogale+Neomys,* from 0.69 to 0.93; *Soriculus+Nesiotites+Chodsigoa+Episoriculus* (excluding *E. fumidis*), from 0.79 to 0.92] and allowed us to estimate divergence times among taxa. Our node time estimates are consistently older than those suggested by He et al. (2010), although Highest Posterior Densities overlap substantially for comparable nodes. For example, the crown of Nectogalini was at 8.5 Ma (95 % highest posterior density, HPD= 6.36–10.78) in our tree versus 6.63 Ma (95 % CI= 4.81–8.84) in He et al. (2010), but around 11 Ma in Dubey et al. (2007). The split between *Chimarrogale* and *Nectogale* genera was 4.83 Ma (95 % HDP= 3.2–6.49) in our tree and 3.71 (95 % CI= 2.1–5.46) in He et al. (2010). Our BEAST analysis suggests a divergence time of 6.44 Ma (95 % HPD=4.68–8.36 Ma) between *N. hidalgo* and *S. nigrescens*.

4. Discussion

Phylogenetic analyses of our molecular data provided strong support for a close relationship between *N. hidalgo* and the Himalayan shrew (*S. nigrescens*), unequivocally refuting suggestions of an affinity between the Mediterranean island shrews and the Asian water shrews *Chimarrogale* and *Nectogale*. Instead, *Chimarrogale* and *Nectogale* form a weakly supported clade with a genus of Eurasian water shrews, *Neomys*, consistent with the results of previous analyses of mitochondrial data (*e.g.*, He et al., 2010; Ohdachi et al. 2006). A monophyletic clade of semi-aquatic shrews is also supported by recent studies based on chromosomal staining, which show that *Neomys* and *Chimarrogale* share karyotype features not present in extant *Soriculus*, *Chodsigoa*, and *Episoriculus* (Mori et al. 2016; Motokawa et al. 2006). However, water shrew monophyly is contradicted by phylogenetic analyses of some nuclear loci (*e.g.*, He et al., 2010), which suggest multiple acquisitions of semi-aquatic adaptations, indicating that further study will be necessary to resolve the issue conclusively. In any case, our molecular phylogenetic analyses suggest that the sister-taxon to *Soriculus+Nesiotites* is actually a clade comprising the other terrestrial nectogaline genera *Chodsigoa* and *Episoriculus* (with the exception of the Taiwanese brown-toothed shrew—*E. fumidus*—which may represent a distinct unrecognised lineage).

Our analyses of combined molecular and morphological data recapitulate the results of past studies that suggested a close relationship between the fossil genus *Asoriculus* and *Nesiotites* (*e.g.*, Alcover et al., 1981; Masini and Sarà, 1998; Maul and Rzebik-Kowalska, 1998; Pons-Monjo et al., 2010; Reumer, 1980, 1984; Rofes and

Cuenca-Bescós, 2006, 2009; Rofes et al. 2012). Our results reinforce the observation of Kormos (1934) concerning the similarities among *A. gibberodon* and *S. nigrescens* (and the Sardinian shrew *Soriculus similis*, which was later ascribed to *Nesiotites* by Bate (1944), although the taxonomy of this latter species remains unclear). This implies that both *Nesiotites* and *Asoriculus* represent a now extinct European radiation of terrestrial nectogaline shrews, with *Nesiotites* as an island form probably derived from mainland *Asoriculus*.

Our molecular dating analyses yielded an estimate of 6.44 Ma (95 % HPD=4.68–8.36 Ma) for the split between *N. hidalgo* and its closest living relative, the Himalayan shrew. This divergence occurred during what appears to be an intense period of nectogaline generic diversification beginning around 8.5 Ma (95 % HPD= 6.36–10.78), which likely contributed to the short internodes and relatively low statistical support for nodes we observed in our time-calibrated phylogeny. Consequently, it seems likely that the divergence between *Nesiotites* and *Asoriculus* occurred soon after the divergence between their common ancestor and the ancestor of the Himalayan shrew (*i.e.*, Late Miocene/Pliocene). This timeline accords well with the hypothesis that the Balearic Islands were colonised by mammal lineages from mainland Europe during the MSC (Bover et al., 2014; Rofes et al., 2012). *A. gibberodon* was widely distributed in Europe from the Late Miocene to Early Pleistocene (*e.g.*, Rofes and Cuenca-Bescós, 2006; Rzebik-Kowalska, 2013), and it has been considered as an indicator of humid environments or even watercourses (*e.g.*, Rofes and Cuenca-Bescós, 2006). Thus, it seems plausible that the common ancestor of *Asoriculus* and *Nesiotites* reached the Balearic Islands from mainland Europe via streams or humid passages, as must also have been the case for other taxa with similar ecological demands (*e.g.*, amphibians; Torres-Roig et al., 2017).

Ultimately, our results support the idea that *Nesiotites* is an insular descendent of a European radiation of terrestrial nectogaline shrews, which became isolated on Mallorca during the MSC. However, additional finds from the depauperate fossil record of Soricinae from the Late Miocene/Early Pliocene from Eurasia (Rzebik-Kowalska, 1998, 2003; Storch et al., 1998) are needed to shed further light on the still confusing taxonomy and paleobiogeography of this clade. The recovery of mitochondrial and nuclear sequences from additional living and extinct soricines will also add important details to our understanding of shrew evolution and biogeography. In this sense, our study is especially important in highlighting the possibilities of retrieving genetic information, even whole mitochondrial genomes, from extinct small mammals (see also, as example, Brace et al., 2016), even from marginal environments for DNA preservation such as the Balearic Islands.

Acknowledgements

Thanks are due to Violaine Nicolas and the Vertebrate Collections of the Muséum National d'Histoire Naturelle of Paris (France) for allowing the morphological character analysis of specimens of *Anurosorex squamipes*, *Soriculus nigrescens*, *Nectogale elegans*, and *Chimarrogale himalayica*.

This work was supported by a Marie Curie International Outgoing Fellowship within the 7th European Community Framework Programme (project MEDITADNA, PIOF-GA-2011-300854, FP7-PEOPLE) (Pere Bover) and a Marie Curie Fellowship (MCA-IEF PIEF-GA-2013-629604) (Juan Rofes). Gloria Cuenca and Juan Rofes are members of the Atapuerca Project CGL2015-65387-C3-2-P (MINECO/FEDER).

The research is also included in Research Project CGL2016-79795, ["Cambios holocénicos en la biodiversidad animal de las islas de la Macaronesia y de Baleares. II" (Dirección General de Investigación Científica y Técnica, Ministerio de Economía y Competitividad)].

K.J.M, B.L., V.T., and A.C. are supported by the Australian Research Council.

Grid computing facilities were provided by CIPRES (Cyberinfrastructure for Phylogenetic Research; https://www.phylo.org).

References

Alcover, J.A., Moyà-Solà, S., Pons-Moyà, J., 1981. Les Quimeres del Passat. Els vertebrats fòssils del Plio-Quaternari de les Balears i Pitiüses. Monografies Científiques 1. Editorial Moll, Palma de Mallorca, Spain.

Altschul, S., Gish, W., Miller, W., Myers, E., Lipman, D., 1990. Basic local alignment search tool. J. Mol. Biol. 215, 403–410.

Arntzen, J.W., García-París, M., 1995. Morphological and allozyme studies of midwife toads (genus *Alytes),* including the description of two new taxa from Spain. Contrib. Zool. 65, 5–34.

Bate, DMA., 1944. Pleistocene shrews from the larger Western Mediterranean Islands. Annals Mag. Nat. Hist. Series 11, 738–769.

Bernhart, S.H., Hofacker, I.L., Will, S., Gruber, A.R., Stadler, P.F., 2008. RNAalifold: improved consensus structure prediction for RNA alignments. BMC Bioinformatics 9, 474.

Bover, P., Alcover, J.A., 2003. Understanding Late Quaternary extinctions: the case of *Myotragus balearicus* (Bate, 1909). J. Biogeogr. 30, 771–781.

Bover, P., Alcover, J.A., 2008. Extinction of the autochthonous small mammals of Mallorca (Gymnesic Islands, Western Mediterranean) and its ecological consequences. J. Biogeogr. 35, 1112–1122.

Bover, P., Rofes, J., Bailón, S., Agustí, J., Cuenca-Bescós, G., Torres, E., Alcover, J.A., 2014. The Late Miocene/Early Pliocene vertebrate fauna from Mallorca (Balearic Islands, Western Mediterranean): an update. Integr. Zool. 9, 183–196.

Bover, P., Valenzuela, A., Torres, E., Cooper, A., Pons, J., Alcover, J.A., 2016. Closing the gap: new data on the last documented *Myotragus* and the first human evidence on Mallorca (Balearic Islands, Western Mediterranean Sea). Holocene 26, 1887–1891.

Brace, S., Thomas, J.A., Dalén, L., Burger, J., MacPhee, R.D., Barnes, I., Turvey, S.T., 2016. Evolutionary history of the Nesophontidae, the last unplaced recent mammal family. Mol. Biol. Evol. 33, 3095–3103.

Briggs, A.W., Stenzel, U., Johnson, P.L.F., Green, R.E., Kelso, J., Prüfer, K., Meyer, M., Krause, J., Ronan, M.T., Lachmann, M., Pääbo, S., 2007. Patterns of damage in genomic DNA sequences from a Neandertal. Proc. Natl. Acad. Sci. U. S. A. 104, 14616–14621.

Brotherton, P., Haak, W., Templeton, J., Brandt, G., Soubrier, J., Adler, C.J., Richards, S.M., Der Sarkissian, C., Ganslmeier, R., Friederich, S., Dresely, V., van Oven, M., Kenyon, R., Van der Hoek, M.B., Korlach, J., Luong, K., Ho, S.Y.W., Quintana-Murci, L., Behar, D.M., Meller, H., Alt, K.W., Cooper, A., the Genographic Consortium, 2013. Neolithic mitochondrial haplogroup H genomes and the genetic origin of Europeans. Nat. Commun. 4, 1764.

Chevreux, B., Wetter, T., Suhai, S., 1999. Genome sequence assembly using trace signals and aditional sequence information, in: Wingender E. (Ed.), Computer Science and Biology: Proceedings of the German Conference on Bioinformatics (GCB). GBF-Braunschweig, Hannover, pp. 45–56.

Chueca, L.J., Gómez-Moliner, B.J., Forés, M., Madeira, M.J., 2017. Biogeography and radiation of the land snail genus *Xerocrassa* (Geomitridae) in the Balearic Islands. J. Biogeogr. 44, 760–772.

Drummond, A.J., Suchard, M.A., Xie, D., Rambaut, A., 2012. Bayesian phylogenetics with BEAUti and the BEAST 1.7. Mol. Biol. Evol. 29, 1969–1973.

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

Fromhage, L., Vences, M., Veith, M., 2004. Testing alternative vicariance scenarios in Western Mediterranean discoglossid frogs. Mol. Phylogenet. Evol. 31, 308–322.

Fu, C., Wei, H., Zong, H., Chen, G., Wang, Q., Yong, B., Chen, S., 2016. The complete mitogenome of Hodgson's Red-Toothed Shrew. *Episoriculus caudatus* (Soricidae). Mitochondrial DNA 27, 1806–1807.

Fumagalli, L., Taberlet, P., Stewart, D.T., Gielly, L., Hausser, J., Vogel, P., 1999. Molecular phylogeny and evolution of *Sorex* shrews (Soricidae: Insectivora) inferred from mitochondrial DNA sequence data. Mol. Phylogenet. Evol. 11, 222–235.

Guimaraes, S., Pruvost, M., Daligault, J., Stoetzel, E., Bennett, E.A., Côté, N.M.-L., Nicolas, V., Lalis, A., Denys, C., Geigl, E.-M., Grange, T., 2017. A cost-effective high-throughput metabarcoding approach powerful enough to genotype ~44000 year-old rodent remains from Northern Africa. Mol. Ecol. 17, 405–417.

Hahn, C., Bachmann, L., Chevreux, B., 2013. Reconstructing mitochondrial genomes directly from genomic next-generation sequencing reads – a baiting and iterative mapping approach. Nucleic Acids Res. 41, e129.

He K, Li YJ, Brandley MC, Lin LK, Wang YX, Zhang YP, Jiang XL. 2010. A multi-locus phylogeny of Nectogalini shrews and influences of the paleoclimate on speciation and evolution. Mol. Phylogenet. Evol. 56, 734–746.

Huang, T., Yan, C., Tan, Z., Tu, F., Yue, B., Zhang, X. 2014. Complete mitochondrial genome sequence of *Nectogale elegans*. Mitochondrial DNA 25, 253–254.

Hutterer, R., 1985. Anatomical adaptations of shrews. Mamm. Rev. 15, 43–55.

Hutterer, R., 1994. Generic limits among neomyine and soriculine shrews (Mammalia: Soricidae), in: Neogene and Quaternary mammals of the Paleartic, conference in honour of professor Kazimierz Kowalski. Abstracts. Polish Academy of Sciences, Kraków, pp. 32.

Jónsson, H., Ginolhac, A., Schubert, M., Johnson, P., Orlando, L., 2013. mapDamage 2.0: fast approximate Bayesian estimates of ancient DNA damage parameters. Bioinformatics 29, 1682–1684.

Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., Buxton, S., Cooper, A., Markowitz, S., Duran, C., Thierer, T., Ashton, B., Mentjies, P., Drummond, A., 2012. Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics 28, 1647–1649.

Kircher, M., 2012. Analysis of high-throughput ancient DNA sequencing data, in: Shapiro, B., Hofreiter, M. (Eds.), Ancient DNA: methods and protocols. Methods Mol. Biol. 840, 197–228.

Kormos, T., 1934. Neue insektenfresser, fledermäuse und nager aus dem Oberpliozän der Villányer Gegend. Földtani Közlöni 64, 296–321.

Krijgsman, W., Hilgen, F.J., Raffi, I., Sierro, F.J., Wilson, D.S., 1999. Chronology,

causes and progression of the Messinian salinity crisis. Nature 400, 652–655.

Lalueza-Fox, C., Castresana, J., Sampietro, L., Marquès-Bonet, T., Alcover, J.A., Bertranpetit, J., 2005. Molecular dating of caprines using ancient DNA sequences of *Myotragus balearicus* an extinct endemic Balearic mammal. BMC Evol, Biol. 5, 70.

Lalueza-Fox, C., Shapiro, B., Bover, P., Alcover, J.A., Bertranpetit, J., 2002. Molecular phylogeny and evolution of the extinct bovid *Myotragus balearicus*. Mol. Phylogenet. Evol. 25, 501–510.

Lanfear, R., Calcott, B., Ho, S.Y.W., Guindon, S., 2012. PartitionFinder: combined selection of partitioning schemes and substitution models for phylogenetic analyses. Mol. Biol. Evol. 29, 1695–1701.

Leonard, J.A., Shanks, O., Hofreiter, M., Kreuz, E., Hodges, L., Ream, W., Wayne, R.K., Fleischer, R.C., 2007. Animal DNA in PCR reagents plagues ancient DNA research. J. Archaeol. Sci. 34, 1361–1366.

Lewis, P.O., 2001. A likelihood approach to estimating phylogeny from discrete morphological character data. Syst. Biol. 50, 913–925.

Li, H., Durbin, R., 2009. Fast and accurate short read alignment with Burrows Wheeler transform. Bioinformatics 25, 1754–1760.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., 1000 Genome Project Data Processing Subgroup, 2009. The Sequence alignment/map (SAM) format and SAMtools. Bioinformatics 25, 2078–2079.

Liu, Z., Zhao, W., Liu, P., Li, S., Xu, C., 2015. The complete mitochondrial genome of Eurasian water shrew (*Neomys fodiens*). Mitochondrial DNA 27, 2381–2382.

Llamas, B., Fehren-Schmitz, L., Valverde, G., Soubrier, J., Mallick, S., Rohland, N., Nordenfelt, S., Valdiosera, C., Richards, S.M., Rohrlach, A., Barreto Romero, M.I., Flores Espinoza, I., Tomasto Cagigao, E., Watson Jiménez, L., Makowski, K., Leboreiro Reyna, I.S., Mansilla Lory, J., Ballivián Torrez, J.A., Rivera, M.A., Burger, R.L., Ceruti, M.C., Reinhard, J., Wells, R.S., Politis, G., Santoro, C.M., Standen, V.G., Smith, C., Reich, D., Ho, S.Y.W. Cooper, A., Haak, W., 2016. Ancient mitocondrial DNA provides high-resolution time scale of the peopling of the Americas. Sci. Adv. 2, e1501385.

Macqueen, P., Seddon, J.M., Austin, J.J., Hamilton, S., Goldizen, A.W., 2010. Phylogenetics of the pademelons (Macropodidae: Thylogale) and historical biogeography of the Australo-Papuan region. Mol. Phylogenet. Evol. 57, 1134–1148.

Manzi, V., Gennari, R., Hilgen, F., Krijgsman, W., Lugli, S., Roveri, M., Sierro, F.J., 2013. Age refinement of the Messinian salinity crisis onset in the Mediterranean. Terra Nova 25, 315–327.

Masini, F., Sarà, M., 1998. *Asoriculus burgioi* sp. nov. (Soricidae, Mammalia) from the Monte Pellegrino faunal complex (Sicily). Acta Zool. Cracov. 41, 111–124.

Maul, L., Rzebik-Kowalska, B., 1998. A record of *Macroneomys brachygnathus* Fejfar, 1966 (Mammalia, Insectivora, Soricidae) in the early Middle Pleistocene (late Biharian) locality of Voigtstedt (Germany) and the history of genus *Macroneomys*. Acta Zool. Cracov. 41, 79–100.

Meyer, M., Kircher, M., 2010. Illumina sequencing library preparation for highly multiplexed target capture and sequencing. Cold Spring Harb. Protoc. 6, 1–10.

Mitchell, K.J., Bray, S.C., Bover, P., Soibelzon, L., Schubert, B.W., Prevosti, F., Prieto, A., Martin, F., Austin, J.J., Cooper, A., 2016. Ancient mitochondrial DNA reveals convergent evolution of giant short-faced bears (Tremarctinae) in North and South America. Biol. Lett. 12, 20160062.

Mora, E., Paspati, A., Decae, A.E., Arnedo, M.A., 2017. Rafting spiders or drifting islands? Origins and diversification of the endemic trap-door spiders from the Balearic Islands, Western Mediterranean. J. Biogeogr. 44, 924–936.

Mori, A., Obara, Y., Kawada, S., Vogel, P., 2016. Chromosomal relationships between two species of water shrew, *Chimarrogale platycephalus* and *Neomys fodiens*. Mammal. Study 41, 17–23.

Motokawa, M., Harada, M., Apin, L., Yasuma, S., Yuan, S.L., Lin, L.K., 2006. Taxonomic study of the water shrews *Chimarrogale himalayica* and *C. platycephala*. Acta Theriol. 51, 215–223.

Ohdachi, S.D., Hasegawa, M., Iwasa, M.A., Vogel, P., Oshida, T., Lin, L.K., Abe, H., 2006. Molecular phylogenetics of soricid shrews (Mammalia) based on mitochondrial cytochrome b gene sequences: with special reference to the Soricinae. J. Zool. 2070, 177–191.

Pons-Monjo, G., Moyà-Solà, S., Furió, M., 2010. El género *Nesiotites* (Soricidae, Insectivora, Mammalia) en las Islas Baleares: estado de la cuestión, in: Moreno-Azanza, M., Díaz-Martínez, I., Gasca, J.M., Melero-Rubio, M., Rabal-Garcés, R., Sauqué, V. (Eds.), VIII Encuentro de Jóvenes Investigadores en Paleontología, volumen de actas. Cidaris 30, 253–258.

Pons-Monjo, G., Moyà-Solà, S., Furió, M., 2012. New data on the origin of *Nesiotites* (Soricidae, Mammalia) in Menorca (Balearic Islands, Spain). C. R. Palevol. 11, 393–401.

Pons-Moyà, J., 1990. Estratigrafía y fauna del yacimiento kárstico de Cala Morlanda (Manacor, Mallorca). Endins 16, 59–62.

Quérouil, S., Hutterer, R., Barrière, P., Colyn, M., Peterhans, J.C.K., Verheyen, E., 2001. Phylogeny and evolution of African shrews (Mammalia: Soricidae) inferred from 16s rRNA sequences. Mol. Phylogenet. Evol. 20, 185–195.

Rambaut, A., Drummond, A.J., 2010. TreeAnnotator version 1.6.1 [software]. http://beast.bio.ed.ac.uk.

Reumer, J.W.F., 1979. On two new micromammals from the Pleistocene of Mallorca. Proc. Konink. Ned. Akad. Wetenschappen B 82, 473–482.

Reumer, J.W.F., 1981. The Pleistocene small mammals from Sa Pedrera de s'Ònix, Majorca (Gliridae, Soricidae). Proc. Konink. Ned. Akad. Wetenschappen B 84, 3–11.

Reumer, J.W.F., 1980. On the Pleistocene shrew *Nesiotites hidalgo* Bate, 1944 from Majorca (Soricidae, Insectivore). Proc. Konink. Ned. Akad. Wetenschappen B 83, 38–68.

Reumer, J.W.F., 1984. Ruscinian and early Pleistocene Soricidae. Geol. Surv. Prof. Paper 565, 1–74.

Reumer, J.W.F., 1989. Speciation and evolution in the Soricidae (Mammalia: Insectivora) in relation with the paleoclimate. Rev. Suisse Zool. 96, 81–90.

Reumer, J.W.F., 1994. Phylogeny and distribution of the Crocidosoricinae (Mammalia: Soricidae), in: Merritt, J.F., Kirkland, G.L.Jr., Rose, R.K. (Eds.), Advances in the biology of shrews. Carnegie Museum Natural History Special Publication, Pittsburgh, pp. 345–356.

Reumer, J.W.F., 1998. A classification of the fossil and recent shrews, in: Wójcik, J.M., Wolsan, M. (Eds.), Evolution of shrews. Polish Academy of Sciences, Bialowieza, pp. 5–22.

Rofes, J., Bover, P., Cuenca-Bescós, G., Alcover, J.A., 2012. *Nesiotites rafelinensis* sp. nov., the earliest shrew (Mammalia, Soricidae) from the Balearic Islands, Spain. Palaeontol. Electronica 15, 8A.

Rofes, J., Cuenca-Bescós, G., 2006. First evidence of the Soricidae (Mammalia) *Asoriculus gibberodon* (Petényi, 1864) in the Pleistocene of North Iberia. Riv. Ital. Paleontol. Strat. 112, 301–315.

Rofes, J., Cuenca-Bescós, G., 2009. A new genus of red-toothed shrew (Mammalia, Soricidae) from the Early Pleistocene of Gran Dolina (Atapuerca, Burgos, Spain), and a phylogenetic approach to the Eurasiatic Soricinae. Zool. J. Linn. Soc. 155, 904–925.

Ronquist, F., Teslenko, M., van der Mark, P., Ayres, D.L., Darling, A., Höhna, S., Larget, B., Liu, L., Suchard, M.A., Huelsenbeck, J.P., 2012. Mrbayes 3.2: Efficient bayesian phylogenetic inference and model choice across a large model space. Syst. Biol. 61, 539–542.

Rzebik-Kowalska, B., 1998. Fossil history of shrews in Europe, in: Wójcik, J.M., Wolsan, M. (Eds.), Evolution of shrews. Polish Academy of Sciences, Bialowieza, pp. 23–92.

Rzebik-Kowalska, B., 2003. Distribution of shrews (Insectivora, Mammalia) in time and space. Deinsea 10, 499–508.

Rzebik-Kowalska, B., 2013. *Sorex bifidus* n. sp. and the rich insectivore mammal fauna (Erinaceomorpha, Soricomorpha, Mammalia) from the Early Pleistocene of Zabia Cave in Poland. Palaeontol. Electronica 16, 12A.

Schubert, M., Lindgreen, S., Orlando, L., 2016. AdapterRemoval v2: rapid adapter trimming, identification, and read merging. BMC Res. Notes. 12, 88.

Storch G, Qiu Z, Zazhigin VS. 1998. Fossil history of shrews in Asia, in: Wójcik, J.M., Wolsan, M. (Eds.), Evolution of shrews. Polish Academy of Sciences, Bialowieza, pp. 93–120.

Torres-Roig, E., Bailon, S., Bover, P., Alcover, J.A., 2017. An early Pliocene anuran assemblage from Mallorca (Balearic Islands, Western Mediterranean): dispersal and palaeobiogeographic implications. Paleobiodivers. Paleoenviron. 97, 315–327.

Willows-Munro, S., Matthee, C.A., 2011. Exploring the diversity and molecular evolution of shrews (family Soricidae) using mtDNA cytochrome b data. Afr. Zool. 46, 246–262.

Wilson, D.E., Reeder, D.M., 2005. Mammal species of the World. A taxonomic and geographic reference. 3rd ed. Johns Hopkins University Press, Baltimore, USA.

Fig. 1. (A) Location of the Mallorcan caves from where samples were used in this paper; (B, C, and D) Lingual view of *Nesiotites hidalgo* mandibles extracted for this study. (B) Pool A, all left mandibles: ACAD13280 (top), ACAD13281 (centre), ACAD13285 (bottom); (C) Pool B: left mandible ACAD13286 (top), right mandible ACAD13289 (centre), and right mandible ACAD13292 (bottom); (D) ACAD14872, right mandible. Scale bar 5 mm.

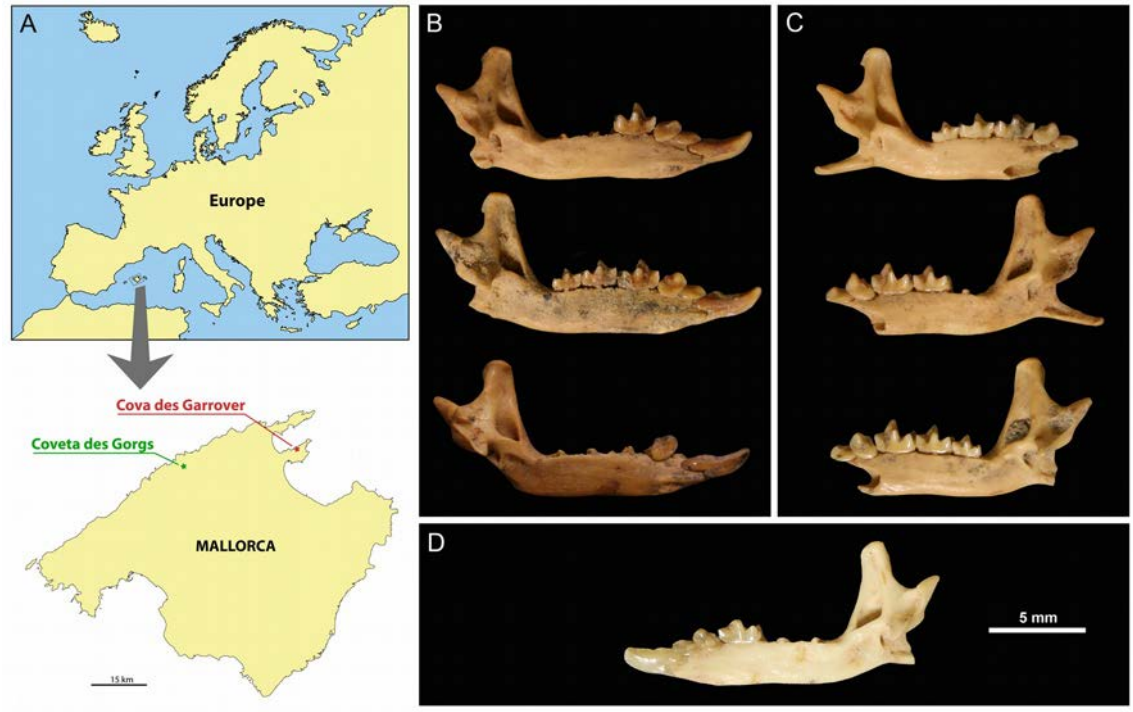


Fig. 2. Combined Bayesian tree using molecular (15,167 bp of the mitochondrial genome) and morphological [48 morphological characters used in Rofes and Cuenca-Bescós (2009)] data. Shadowed names indicate fossil species. Asterisks in nodes indicate a posterior probability (PP) = 1. See Table S4 for further information on data used.

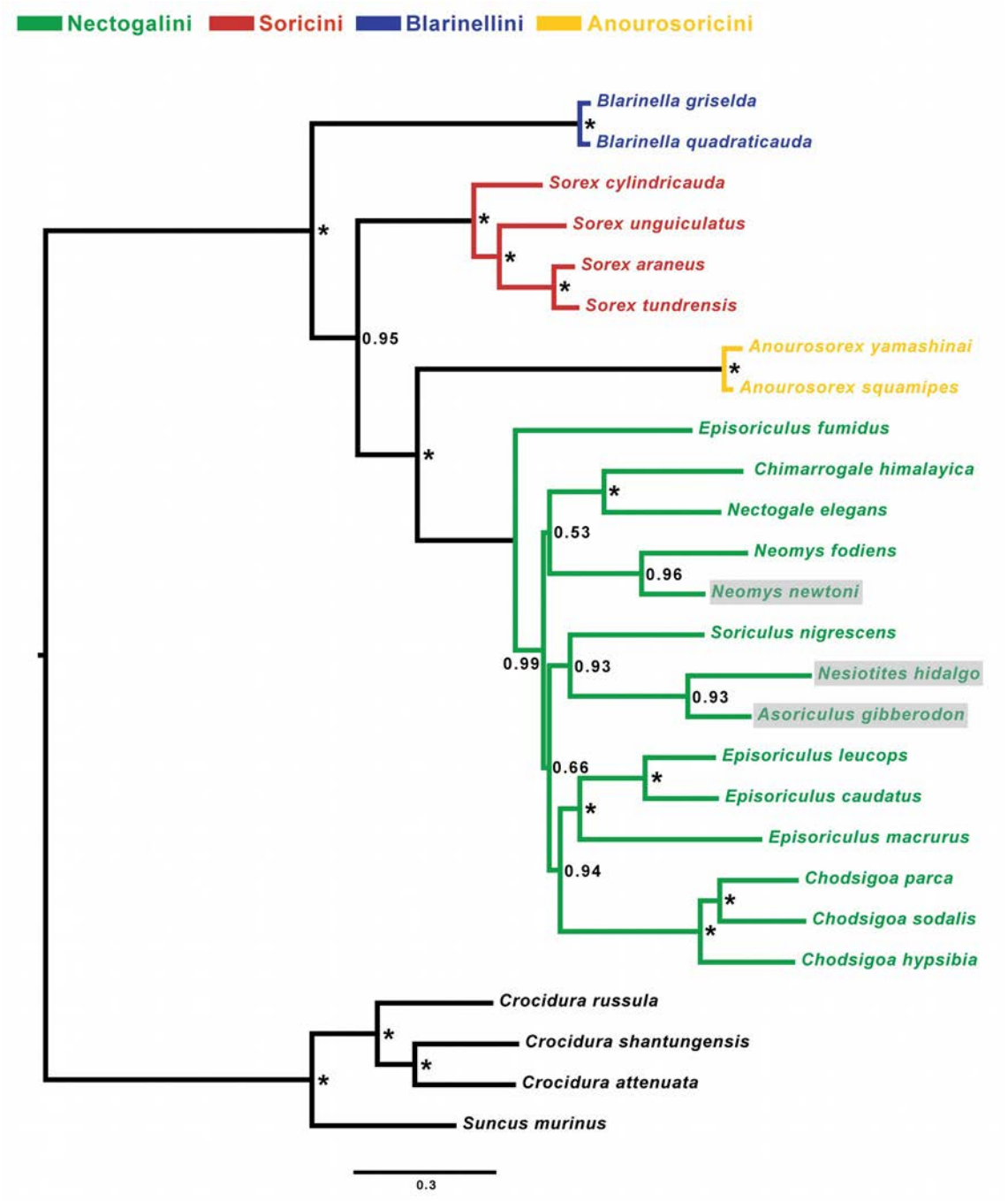
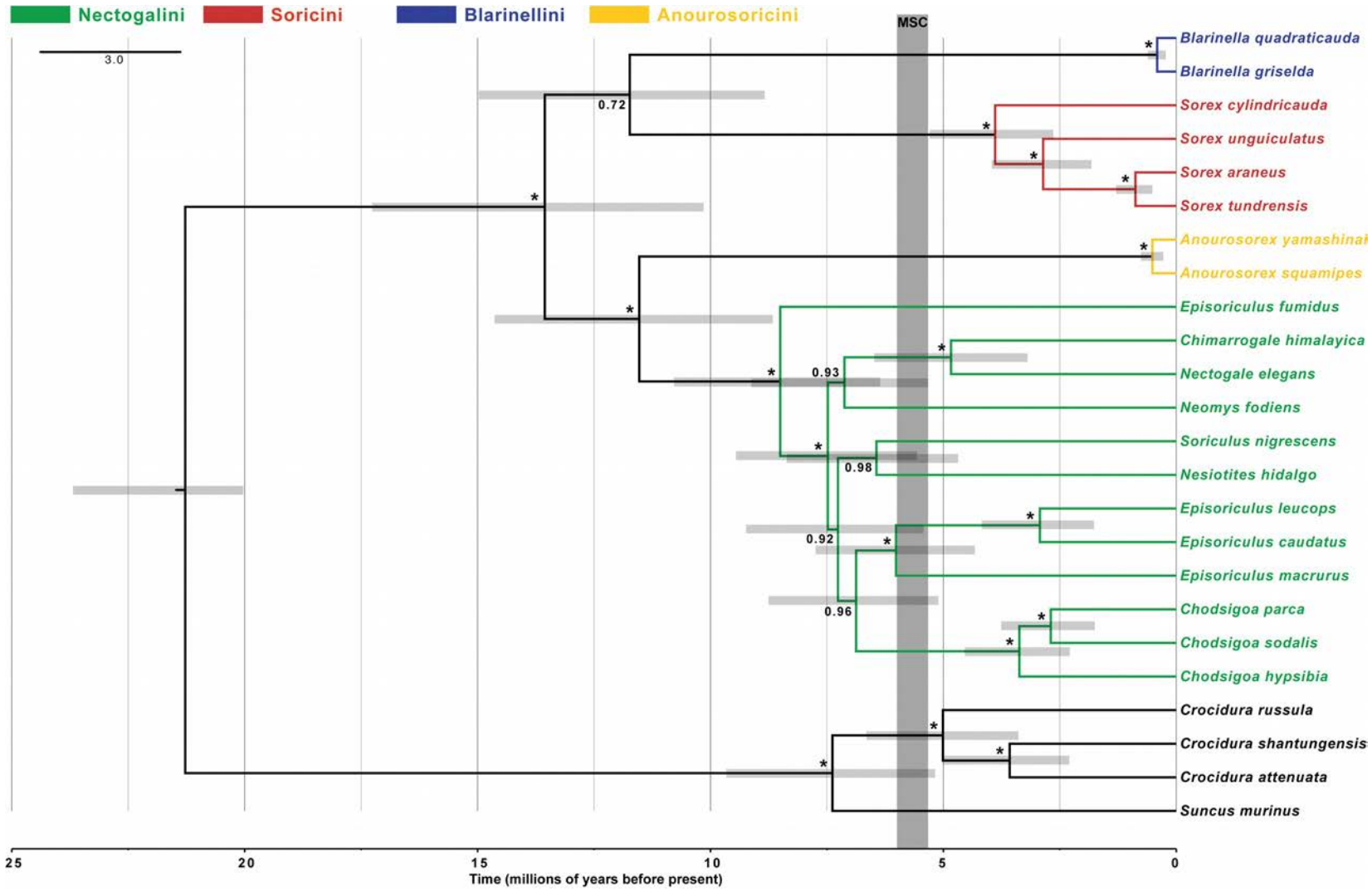


Fig. 3. Calibrated Bayesian tree of Soricinae using split Soricinae-Crocidurinae as calibration age. Branch support (posterior probability, PP) are given for each clade and nodes with asterisks indicate a posterior probability (PP) = 1. Grey bars reflect the 95 % highest posterior densities (HPDs) of each node. Dark shadowed area indicates the chronology for the Messinian Salinity Crisis (MSC).



Supplementary Tables

Table S1. Radiocarbon dates of fossil vertebrates obtained from the same caves from where the *Nesiotites hidalgo* samples used in this paper were obtained.

Lab code	Species	Deposit	Radiocarbon Age (BP)	2σ Calibration interval (cal BC/AD)	Source
RICH-21771	*Myotragus balearicus*	Coveta des Gorgs	4456±33	3340–3010	Bover et al., 2016
RICH-21776	*Myotragus balearicus*	Coveta des Gorgs	5438±35	4350–4230	Bover et al., 2016
RICH-21975	*Myotragus balearicus*	Coveta des Gorgs	6313±37	5370–5210	Bover et al., 2016
RICH-21769	*Myotragus balearicus*	Coveta des Gorgs	6588±35	5620–5580 (17.1 %) 5570–5480 (78.3 %)	Bover et al., 2016
RICH-21768	*Myotragus balearicus*	Coveta des Gorgs	6594±35	5620–5480	Bover et al., 2016
Beta-177239	*Myotragus balearicus*	Coveta des Gorgs	7060±40	6016–5847	Lalueza-Fox et al., 2005
KIA-29167	*Myotragus balearicus*	Coveta des Gorgs	7580±35	6476–6396	This paper
KIA-29176	*Myotragus balearicus*	Coveta des Gorgs	8390±40	7542–7354	This paper
RICH-21395	*Myotragus balearicus*	Coveta des Gorgs	8372±40	7531–7346	This paper
RICH-21773	*Myotragus balearicus*	Coveta des Gorgs	8562±38	7610–7530	This paper
Beta-143117	*Myotragus balearicus*	Coveta des Gorgs	8570±40	7651–7534	Bover and Alcover, 2003
RICH-21974	*Myotragus balearicus*	Coveta des Gorgs	9164±42	8540–8510 (2.4 %) 8480–8280 (93 %)	This paper
Beta-163133	*Nesiotites hidalgo*	Cova des Garrover	4280±50	3082–2698	Bover and Alcover, 200

Table S2. Soricidae species and accesion numbers used in the phylogenetic analyses. Individuals with just one GenBank accession number indicates whole mitocondrial genomes

	CYB	**CO1**	**ATP6**	**ND2**	**ND4**	**ND5**	**12S_rRNA**	**16S_rRNA**
Anourosorex squamipes	KJ545899							
Anourosorex yamashinai	GU981257	GU981211	GU981136	GU981303	GU981349	GU981395	GU981015	GU981060
Blarinella griselda	GU981258	GU981212	GU981137	GU981304	GU981350	GU981396	GU981016	GU981067
Blarinella quadraticauda	KJ131179							
Chimarrogale himalayica	GU981263	GU981217	GU981142	GU981309	GU981355	GU981401	GU981021	GU981061
Chodsigoa hypsibia	GU981260	GU981214	GU981139	GU981306	GU981352	GU981398	GU981018	GU981069
Chodsigoa parca	GU981265	GU981219	GU981144	GU981311	GU981357	GU981403	GU981023	GU981063
Chodsigoa sodalis	GU981269	GU981223	GU981148	GU981315	GU981361	GU981407	GU981027	GU981075
Crocidura attenuata	KP120863							
Crocidura russula	NC_006893							
Crocidura shantungensis	JX968507							
Episoriculus caudatus	KM503097							
Episoriculus fumidus	NC_003040							
Episoriculus leucops	GU981281	GU981235	GU981160	GU981327	GU981373	GU981419	GU981039	GU981086
Episoriculus macrurus	KU246040							
Nectogale elegans	KC503902							
Neomys fodiens	KM092492							
Sorex araneus	KT210896							
Sorex cylindricauda	KF696672							
Sorex tundrensis	KM067275							
Sorex unguiculatus	NC_005435							
Soriculus nigrescens	GU981297	GU981251	GU981176	GU981343	GU981389	GU981435	GU981055	GU981102
Suncus murinus	KJ920198							

Table S3. Best partition schemes for MrBayes and Beast analyses inferred by PartitionFinder.

MrBAYES		
Partition	**Best Model**	**Subset Sites**
1	GTR+I+G	ND1_1, ND2_1, ND3_1, ND4L_1, ND4_1, ND5_1
2	SYM+I+G	ATP6_1, CO1_1, CO2_1, CO3_1, CYB_1
3	GTR+I+G	12S_rRNA_loops, 16S_rRNA_loops, ATP8_1, tRNA_loops
4	GTR+I+G	ATP6_2, CO2_2, CO3_2, CYB_2, ND1_2, ND2_2, ND3_2, ND4L_2, ND4_2, ND5_2
5	F81+I	CO1_2
6	HKY+I+G	ATP8_2, ND6_1, ND6_2
7	GTR+I+G	12S_rRNA_stems, 16S_rRNA_stems, tRNA_stems
8	HKY+I+G	CO3_3, ND1_3, ND3_3, ND4L_3, ND4_3, ND5_3, ND6_3
9	HKY+I+G	CYB_3, ND2_3
10	GTR+I+G	ATP6_3, ATP8_3, CO1_3, CO2_3

BEAST		
Partition	**Best Model**	**Subset Sites**
1	GTR+I+G	ATP8_2, ND1_1, ND2_1, ND3_1, ND4L_1, ND4_1, ND5_1, ND6_1
2	SYM+I+G	ATP6_1, CO1_1, CO2_1, CO3_1, CYB_1
3	GTR+I+G	12S_rRNA_loops, 16S_rRNA_loops, ATP8_1, tRNA_loops
4	TVM+I+G	ATP6_2, ND1_2, ND2_2, ND3_2, ND4L_2, ND4_2, ND5_2
5	K81uf+I+G	CO1_2, CO2_2, CO3_2, CYB_2
6	TVM+I+G	12S_rRNA_stems, 16S_rRNA_stems, ND6_2, tRNA_stems
7	TrN+I+G	CO3_3, ND1_3, ND2_3, ND3_3, ND4L_3, ND4_3, ND5_3, ND6_3
8	TrN+I+G	ATP6_3, ATP8_3, CO1_3, CO2_3
9	HKY+I+G	CYB_3

Table S4. Data availability for combined morphological/molecular phylogeny. Data type available (marked in grey) for the combined morphological/molecular Bayesian phylogenetic tree (Fig. 2) using morphological characters following Rofes and Cuenca-Bescós (2009) and molecular data from Genbank (Table S2) and *Nesiotites* mitocondrial genome data. Names in red indicates extinct taxa.

Species	Morphological data	Molecular data
Anourosorex squamipes		
Anourosorex yamashinai		
Asoriculus gibberodon		
Blarinella Griselda		
Blarinella quadraticauda		
Chimarrogale himalayica		
Chodsigoa hypsibia		
Chodsigoa parca		
Chodsigoa sodalis		
Crocidura attenuata		
Crocidura fuliginosa		
Crocidura shantungensis		
Episoriculus caudatus		
Episoriculus fumidus		
Episoriculus leucops		
Episoriculus macrurus		
Nectogale elegans		
Neomys fodiens		
Neomys newtoni		
Nesiotites hidalgo		
Sorex bedfordiae		
Sorex cylindricauda		
Sorex tundrensis		
Sorex unguiculatus		
Soriculus nigrescens		
Suncus murinus		

Supplementary figures

Fig. S1. MapDamage report of the BWA mapping to the *Nesiotites hidalgo* mitochondrial genome consensus generated using MITObim. Substitution panels (left) show the characteristic high frequency of purines and pyrimidines just before and after the reads, respectively, and the accumulation of 5' C-to-T (red) and 3' G-to-A (blue) misincorporations characteristic of ancient DNA. The truncated pattern of the curve observed at the 3' extreme is the result of the trimming of the first nucleotide of all reverse reads (due to the presence of an A in the first position of all these reads). On the right, fragment size distribution of the mapped unique reads.

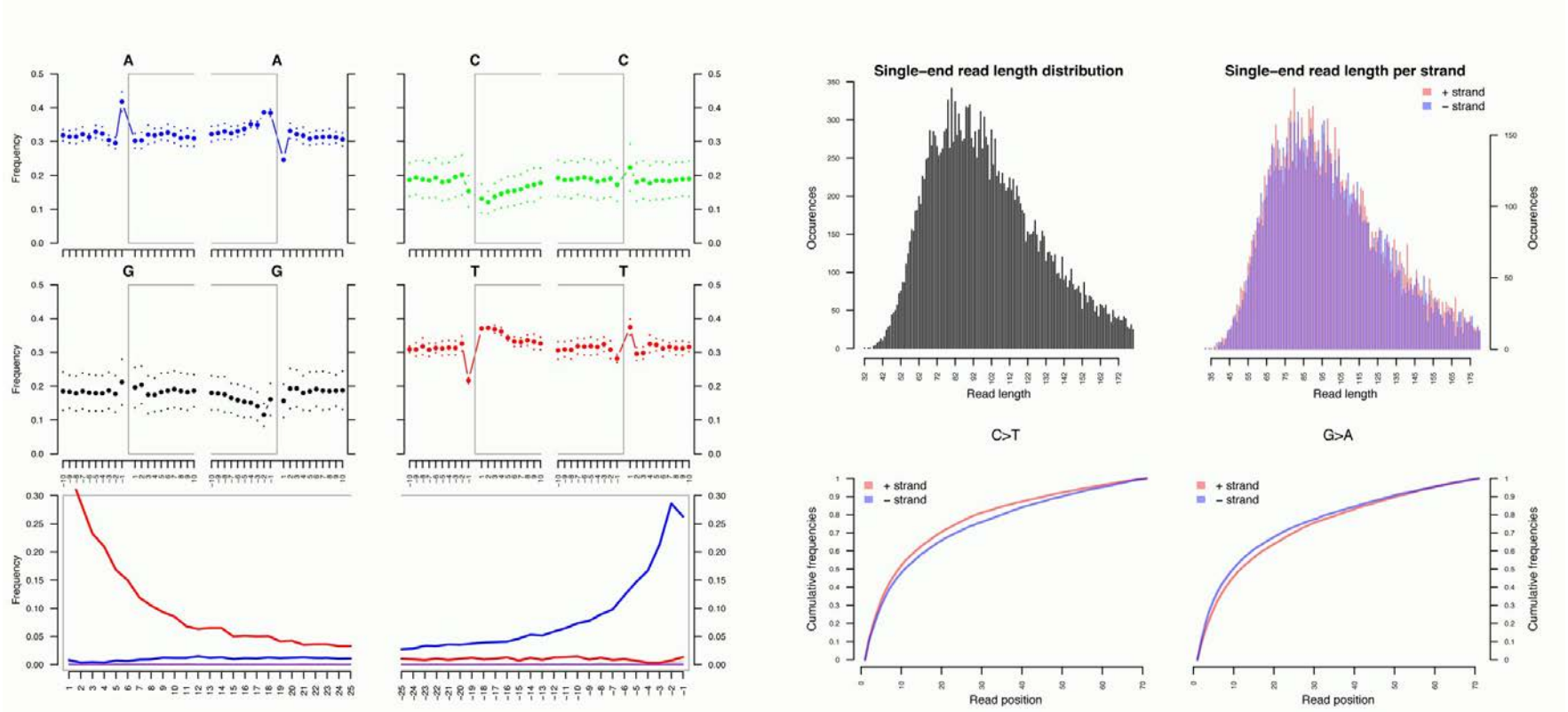


Fig. S2. Bayesian phylogeny. Phylogenetic relationships in Soricinae (Crocidurinae as outgroup) in MrBayes using 15,167 bp of the mitocondrial genome. Asterisks in nodes indicate a posterior probability (PP) = 1.