4-20-2016

Mitogenomic sequences support a north–south subspecies subdivision within Solenodon paradoxus

Adam L. Brandt

Kirill Grigorev

Yashira M. Afanador-Hernández

Liz A. Paullino

William J. Murphy

See next page for additional authors

Authors

Adam L. Brandt, Kirill Grigorev, Yashira M. Afanador-Hernández, Liz A. Paullino, William J. Murphy, Adrell Núñez, Aleksey Komissarov, Jessica R. Brandt, Pavel Dobrynin, David Hernández-Martich, Roberto María, Stephen J. O'Brien, Luis E. Rodríguez, Juan C. Martínez-Cruzado, Taras K. Oleksyk, and Alfred L. Roca

Mitochondrial DNA, Original Article

Title: Mitogenomic sequences support a north-south subspecies subdivision within *Solenodon paradoxus*

Authors: Adam L. Brandt[1,2+], Kirill Grigorev[4+], Yashira M. Afanador-Hernández[4], Liz A. Paulino[5], William J. Murphy[6], Adrell Núñez[7], Aleksey Komissarov[8], Jessica R. Brandt[1], Pavel Dobrynin[8], J. David Hernández-Martich[9], Roberto María[7], Stephen J. O'Brien[8,10], Luis E. Rodríguez[5], Juan C. Martínez-Cruzado[4], Taras K. Oleksyk[4*] and Alfred L. Roca[1,2,3*]

[+]Equal contributors
*Corresponding authors: roca@illinois.edu; taras.oleksyk@upr.edu

Affiliations: [1]Department of Animal Sciences, [2]Illinois Natural History Survey and [3]Carl R. Woese Institute for Genomic Biology, University of Illinois at Urbana-Champaign, Urbana, IL 61801; [4]Department of Biology, University of Puerto Rico at Mayagüez, Mayagüez, Puerto Rico 00681; [5]Instituto Tecnológico de Santo Domingo (INTEC), Santo Domingo, Dominican Republic; [6]Department of Veterinary Integrative Biosciences, Texas A&M University, College Station, TX 77843; [7]Department of Conservation and Science, Parque Zoológico Nacional (ZOODOM), Santo Domingo, Dominican Republic; [8]Theodosius Dobzhansky Center for Genome Bioinformatics, St. Petersburg State University, 199004 St. Petersburg, Russia; [9]Instituto de Investigaciones Botanicas y Zoologicas, Universidad Autónoma de Santo Domingo, Santo Domingo, Dominican Republic; [10]Oceanographic Center, Nova Southeastern University, Fort Lauderdale, Florida 33004

Corresponding address:
Alfred L. Roca, Ph.D.
Department of Animal Sciences
University of Illinois at Urbana-Champaign
432 ASL, 1207 West Gregory Drive
Urbana, Illinois 61801
Email: roca@illinois.edu
Mobile phone: 240-422-2493
Office phone: 217-244-8853
Department fax: 217-333-8286

This is an Accepted Manuscript of an article published by Taylor & Francis Group in Mitochondrial DNA Part A on 15/03/2016, available online: http://dx.doi.org/10.3109/24701394.2016.1167891.

Abstract

Solenodons are insectivores found only in Hispaniola and Cuba, with a Mesozoic divergence date versus extant mainland mammals. Solenodons are the oldest lineage of living eutherian mammal for which a mitogenome sequence has not been reported. We determined complete mitogenome sequences for six Hispaniolan solenodons (*Solenodon paradoxus*) using next generation sequencing. The solenodon mitogenomes were 16,454-16,457 bp long and carried the expected repertoire of genes. A mitogenomic phylogeny confirmed the basal position of solenodons relative to shrews and moles, with solenodon mitogenomes estimated to have diverged from those of other mammals ca. 78 Mya. Control region sequences of solenodons from the northern (n=3) and southern (n=5) Dominican Republic grouped separately in a network, with F_{ST} = 0.72 (p=0.036) between north and south. This regional genetic divergence supports previous morphological and genetic reports recognizing northern (*S. p. paradoxus*) and southern (*S. p. woodi*) subspecies in need of separate conservation plans.

Key words [3-6]: conservation, Dominican Republic, evolutionarily significant units, *Solenodon paradoxus*, subspecies, taxonomy

Introduction

Solenodons are fossorial insectivores that live only in the islands of Cuba and Hispaniola (Ottenwalder 1985). The two species of solenodon are among the few native non-volant mammals to survive human colonization of the West Indies (MacPhee et al. 1999; Ottenwalder 2001). The geographic range of the Cuban solenodon (*Solenodon cubanus* Peters, 1861) is currently limited to eastern Cuba (Ottenwalder 1985). The Hispaniolan solenodon (*S. paradoxus* Brandt, 1833) lives only in the island of Hispaniola, with surviving range primarily within the Dominican Republic (Ottenwalder 1985), but with a small remnant population in Haiti (Turvey et al. 2008). Solenodons inhabit the forests of Cuba and Hispaniola to elevations of 2000 m or more (Eisenberg and Gozalez Gotera 1985; Ottenwalder 1985) and shelter in caves, crevices, logs, and extensive tunnel networks at a depth of >20 cm (Eisenberg and Gozalez Gotera 1985; Ottenwalder 1985). They subsist on small vertebrates and large invertebrates (Eisenberg and Gozalez Gotera 1985; Ottenwalder 1985). Solenodons are one of the few mammals that are venomous (Turvey 2010; Folinsbee 2013), and some have argued that they may use a form of echolocation (Eisenberg and Gould 1966).

Among hundreds of lineages of recent West Indian tetrapods (mammals, birds, reptiles and amphibians), *Solenodon* is the only lineage for which strong evidence supports Mesozoic divergence when compared to the closest related living mainland taxa (Roca et al. 2004; Meredith et al. 2011). Most other lineages in the West Indies are known to have diverged from mainland forms in the Cenozoic (Hedges 1996; Woods and Sergile 2001; Roca et al. 2004). In a previous study using 16 nuclear and 3 mitochondrial

genes, solenodons proved to be the earliest branching lineage within Eulipotyphla, or true insectivores (Roca et al. 2004). Contrary to the expectations of some previous morphological studies, solenodons were basal to soricids (shrews), talpids (moles) and erinaceids (hedgehogs and gymnures) (McDowell 1958; Butler 1988; MacPhee and Novacek 1993; McKenna and Bell 1997). The divergence date for solenodons from other eulipotyphlan insectivores has been estimated as 76 million years ago (Mya; 95% credibility interval [CI] of 72-81 Mya), older than some interordinal splits among mammals (Meredith et al. 2011). This point estimate is before the Cretaceous/Tertiary boundary, with the 95% CI for the solenodon divergence falling entirely within the Mesozoic Era (Hedges 1996; Iturralde-Vinent and MacPhee 1999; Roca et al. 2004; Meredith et al. 2011).

Despite their unique status among mammals (Roca et al. 2004; Wible 2008), until recently only a few studies had published on their DNA, these involving one or a few specimens (Stanhope et al. 1998; Emerson et al. 1999; Allard et al. 2001; Waddell and Shelley 2003; Roca et al. 2004). Likewise, solenodons are the only major mammalian lineage for which the full length of the mitochondrial DNA genome has not been sequenced. Both solenodon species are listed as endangered and declining in population by the IUCN Red List of threatened species (IUCN 2013). Solenodons are currently threatened with extinction due to deforestation, increasing human populations, predation by introduced carnivores, and possibly by competition from introduced rodents (Ottenwalder 1985; Ottenwalder 2001; Borroto-Paez 2009; IUCN 2013; Turvey et al. 2014). A better understanding of the genetics of solenodons would assist development of conservation plans. In particular, regional differences in morphology have been reported

in the Hispaniolan solenodon, with specimens from the north of the island (designated as subspecies *Solenodon paradoxus paradoxus*) larger in size than specimens from the south (designated *S. p. woodi*) (Ottenwalder 2001). We therefore collected Hispaniolan solenodons from the southern Dominican Republic, and compared their sequences to those of northern solenodons. We also generated the first full mitogenomic sequences reported from both subspecies of Hispaniolan solenodon, filling this gap among mammals, and determining whether the mitogenomic analysis would support a Mesozoic divergence date for the lineage.

Materials and Methods

Sample collection and DNA extraction

This study was conducted with the review and approval of the Institutional Animal Care and Use Committee of the University of Puerto Rico. Before the study was conducted, required permits were obtained, including collection and export permits issued by the government of the Dominican Republic, and ESA/CITES and APHIS import permits issued by the US government. For sample Spa-1, collection and DNA extraction details have been reported previously (Roca et al. 2004), all other sequences were from samples newly collected in the wild across various localities in the southern Dominican Republic (Table 1, Figure 1). Samples were obtained in accordance with a permit issued by the Dominican Republic Ministry of the Environment and Natural Resources. As a general protocol, the individuals were captured, weighed, measured,

sexed and immediately released at the capture site, all within ten minutes of capture. A general visual assessment was performed looking for any sign of possible disease. As a general rule, the blood volume collected did not exceed more than 1% of body weight of animals with no visible sign of distress or disease. Aseptic technique was performed previous to blood draw using a povidone-iodine solution, followed by isopropyl alcohol. Whole blood was collected from the *vena jugularis* using a 3 mL syringe with a 23G x 1" needle. Once collected, the samples were transferred to a micro-blood collection tube with anticoagulant (BD Microtainer, 1.0 mg K2EDTA for 250-500µL volume). The specimens were then labeled on their tail with a marker to avoid recapturing. Geographic coordinates were recorded for every location where a specimen was captured (Table 1). Tubes were refrigerated until DNA could be extracted from samples using the DNeasy Blood & Tissue kit (Qiagen, Hilden, Germany).

Mitogenome sequence generation, assembly and annotation

For each of six individuals of *S. paradoxus*, a mitogenome was assembled independently from Illumina WGS reads. From northern solenodon (*S. p. paradoxus*) Spa-1, a total of 52,358,830 paired-end reads were generated, equating to approximately 13.09 Gb of sequence data. For five southern solenodons (*S. p. woodi*), an average of 151,783,327 paired-end reads were generated, equating to an average of 15.33 Gb of sequence data for each individual. For each of the six individuals, a mitogenome was assembled independently using two methods. First via a standard approach with MITO-bim (Hahn et al. 2013) using the complete mitochondrial genome sequence of *Tapirus indicus* (GenBank accession number NC_023838) as a reference seed. Separately, they

were also assembled using a Bruijn graph based algorithm that considers coverage, based on the software Cookiecutter (Starostina et al. 2015). Both produced identical results outside of the control region and open reading frames were present in all coding regions which might otherwise indicate assembly of numts (Lopez et al. 1994). The second software was helpful for excluding reads of unusually low coverage that were almost certainly numts, allowing accurate assembly even for the control region.

Complete mitogenome sequences were assembled separately for each individual. Multiple sequence alignment with MAFFT (Katoh and Standley 2013) and a custom variant calling script were employed to assess variation among mitogenomes. This variation was confirmed by aligning reads from every individual to the assembly of every other individual with Bowtie2 (Langmead and Salzberg 2012) and calling variants with SAMtools and Bcftools (Li 2011). All solenodon mitogenomes reported in this study were annotated and deposited in GenBank (accession numbers KU697358-KU697363). Partial mitogenome sequences previously published for *S. cubanus* (Roca et al. 2004) were likewise aligned to the consensus sequence with MAFFT (Katoh and Standley 2013), and variants were called.

Protein coding gene discovery and annotation for the consensus sequence of *S. paradoxus woodi* was performed with NCBI ORF Finder (Wheeler et al. 2003) and NCBI Blast (Altschul et al. 1990); rRNA gene discovery and annotation were performed with Barrnap (github.com/tseemann/barrnap). The tRNA gene discovery and annotation was performed with tRNAscan SE (Schattner et al. 2005), with 21 of the 22 tRNA coding genes discovered using relaxed models in tRNAscan and Cove for mitochondrial sequences, with a Cove score cutoff of 10 and intermediate score cutoff of 5. The gene

for tRNA-Ser, expected at approximately the 11.5 kilobase mark, was not discovered with this method. We used Exonerate (Slater and Birney 2005) with the ungapped model to align tRNA coding genes of *Sorex araneus* (NCBI accession number NC_027963) to the mitogenome of *Solenodon paradoxus woodi*. All of these aligned to the previously discovered regions, with eight (for Ile, Gln, Ala, Asn, His, Ser, Leu, Pro) alignments spanning their entire lengths, including the previously missed tRNA-Ser at bases 11632–11690. The structure and polymorphisms of the mitogenomes and of mtDNA segments were visualized using the software Circos (Krzywinski et al. 2009).

To assess amino acid substitutions, pairwise alignments of protein coding regions of *S. paradoxus woodi* to *Talpa europaea* (NCBI accession number NC_002391), *Sorex araneus* (NC_027963), *Felis catus* (NC_001700) and *Homo sapiens* (GCA_000001405.20) were performed with the software Exonerate (Slater and Birney 2005) model coding2coding. Provean (Choi and Chan 2015) scores were calculated for the substitutions. A stringent score cutoff of -4.1 was used to identify mutations most likely to have effect on protein function; the complete list of these mutations, grouped by genes, is presented in Supplementary Table S1.

Phylogenetic analyses and molecular dating

For phylogenetic analysis, nucleotide sequences were aligned using webPRANK (Loytynoja and Goldman 2010) in EBI Web Services (McWilliam et al. 2013); alignment output was visually inspected and edited using Mesquite v3.04 (http://mesquiteproject.org). Substitution models were determined using the Akaike information criterion in jModeltest2 v 2.1.8 (Darriba et al. 2012). Model testing and

198 phylogenetic analyses were performed in CIPRES Science Gateway (Miller et al. 2015).
199 Effective sample size values for Markov chain Monte Carlo (MCMC) analyses were
200 verified to be at or above 200 for all parameters. All trees were visualized in FigTree
201 v1.4.2.

202 To examine relationships among the mitogenomes of *S. paradoxus*, the complete
203 mitogenomic alignment (including control region) of 16,457 bp was used for Bayesian
204 phylogenetic inference (BI) implemented in the program BEAST v1.8.2(Drummond et al.
205 2012), using the HKY substitution model, constant coalescent tree prior, and one of two
206 strict molecular clock rates (fast and slow, respectively): 3.72×10^{-8} (*Mus musculus*)
207 (Goios et al. 2007) or 1.665×10^{-8} (*Homo sapiens*) (Soares et al. 2009) substitutions per
208 site per year. Posterior distributions were obtained by MCMC sampling from at least
209 10,000,000 steps, with a discarded burn-in of at least 1,000,000; samples were drawn
210 every 1000 MCMC steps.

211 For another analysis, a 2,495 bp alignment of Cuban solenodon (Roca et al. 2004)
212 sequence spanning the 12S and 16S rRNA regions were aligned to the homologous
213 region of the *S. paradoxus* mitogenomes used to generate a maximum likelihood (ML)
214 tree in RAxML version 8.2.4 (Stamatakis 2014), using the GTR+I+G substitution model
215 and at least 1000 rapid bootstrap replicates. For additional examination of clades that
216 appeared to correspond to northern and southern geographic localities of *S. paradoxus*,
217 three previously published control region sequences (Allard et al. 2001) (two northern
218 and one unknown in origin; Figure 1, Table 1) were aligned to the newly generated
219 mitogenomes. A 389 bp alignment of the control region was used for this analysis.
220 Weighted maximum likelihood distances were calculated and used to generate a median

joining network using the software NETWORK v.4.6.1 (Bandelt et al. 2009). Population differentiation (F_{ST}) between northern and southern solenodons based on 389 bp of the control region was calculated with ARLEQUIN v.3.5 (Excoffier and Lischer 2010).

To place the solenodon into a phylogeny with other mammals, the solenodon mitogenomes (without control region) were aligned to those of the mammals listed in Supplementary Table S2. These included a shrew (Soricidae) and mole (Talpidae), representing the families most closely related to the Solenodontidae, along with 52 additional taxa from Laurasiatheria. The mitogenome of the hedgehog, *Erinaceus europaeus*, was excluded from the analysis due to previously reported (Mouchaty et al. 2000) significant deviation of the published mitogenome sequence from expected nucleotide and amino acid compositions. For this analysis, a total of 15,624 aligned nucleotide positions were analyzed after poorly aligned regions (mainly within 12S and 16S rRNA regions) were excluded. The software RAxML version 8.2.4 (Stamatakis 2014) was used to generate a maximum likelihood tree, using the GTR+I+G substitution model and 1000 rapid bootstrap replicates.

For molecular dating of the split between the Solenodontidae and other laurasiatherian families, fossil calibration dates reported previously (Meredith et al. 2011) were used. These fossil calibration dates are listed in Supplementary Table S3, along with the inferred date estimates for nodes on the mammalian mitogenomic tree. This analysis was performed using BEAST v1.8.2 (Drummond et al. 2012). We utilized an uncorrelated lognormal relaxed molecular clock model, which permits the rate of molecular substitutions to be independent across the tree while incorporating uncertainty in both tree topology and multiple fossil calibrations (Drummond et al. 2012). We used

randomly generated starting trees and birth-death process tree prior (Gernhard 2008). A monophyly constraint was applied to non-eulipotyphlan taxa to minimize spurious dating resulting from incorrect trees and to enforce the correct topology (Eulipotyphla sister to all other laurasiatherians), which has been well established using nuclear data (Roca et al. 2004; Meredith et al. 2011). Posterior distributions were obtained by MCMC sampling from at least 60,000,000 steps, with a discarded burn-in of at least 6,000,000; samples were drawn every 1000 MCMC steps.

Results

The consensus mitogenome of the Hispaniolan solenodon was 16,457 bp, and consists of the expected repertoire of 37 genes, including 13 that code for polypeptides, 22 for transfer RNA (tRNA) and the small and large subunits of ribosomal RNA (rRNA). The length of the mitogenome was: 16,457 bp for Spa-K; 16,456 bp for Spa-L; 16,455 for Spa-M, -N and -O; and 16,454 for Spa-1. Figure 2 represents gene and variant tracks plotted with Circos (Krzywinski 2009). The five individuals of *S. paradoxus woodi* exhibited little sequence variation (Figure 2). Spa-M, -N and -O were identical in sequence. Between Spa-K and Spa-L there were 2 differences. And for Spa-M/N/O there were 4 differences with Spa-K and 2 differences with Spa-L. By contrast, the mitogenome of the individual of *S. paradoxus paradoxus* (Spa-1) exhibited 100 differences from the consensus of the 5 southern solenodons; and the mtDNA fragments of *S. cubanus* showed an even greater density of differences (Figure 2). Coding genes were discovered and annotated *de novo*. The relative positions and lengths of the genes in

S. paradoxus closely corresponded to those in the mitogenomes of other laurasiatherian mammals, corroborating the result. Comparison of the amino acid residues of the protein coding genes with those of four other species of eutherian mammals found many substitutions with significant Provean scores. Mutations with Provean scores indicating mutations most likely to have functional effects were grouped by gene, and are presented in Supplementary Table S1. Most of these radical mutations corresponded to changes that had occurred in the solenodon lineage, since the other species all shared the same amino acid at most positions. Most of the radical mutations fell within the genes *ND2* and *ND5*.

The complete mitogenomes (including control regions) of Hispaniolan solenodons were used to infer a phylogeny in BEAST. The 5 solenodons from southern Hispaniola formed a clade with the northern Hispaniolan solenodon as an outgroup to the clade (Figure 3A). Using two evolutionary rates (the fast rate of house mouse (Goios et al. 2007), 3.72×10^{-8} and the slow rate of humans (Soares et al. 2009), 1.665×10^{-8} substitutions per site per year) as extremes, the point estimate for the divergence between the two clades was estimated as 78,000 and 171,400 years, respectively. Using an alignment of 12S and 16S rRNA regions for which *S. cubanus* sequence could be used as an outgroup confirmed that the split in *S. paradoxus* was between the 5 southern and 1 northern Hispaniolan solendons (Figure 3B). In an attempt to increase the number of northern individuals that could be analyzed, the control region sequences of the *S. paradoxus* genomes were aligned with those previously published for three solenodons, two of northern provenance (Loma de la Jagua; Figure 1) and one of unknown origin. In a network analysis of the control region sequences, there was separation between the northern (n=3) and southern (n=5) individuals (Figure 3C). For this dataset, F_{ST} between

the northern and southern populations was estimated as 0.72477 (p = 0.03604). It is important to note that although F_{ST} between the northern and southern populations was high, there were only a few mutational steps that differentiated the two populations.

Our solenodon mitogenomic sequences (excluding control region) were then aligned with those of 54 other mammalian taxa (Supplementary Table S2). The relationships inferred for the taxa are shown in Supplementary Figures S1 and S2, and confirms that solenodons were basal to soricids and talpids. Relying on well-established fossil calibration dates reported previously (Supplementary Table S3) (Meredith et al. 2011), we estimated that solenodons had diverged from other mammals ca. 78.2 Mya (95% CI of 62.1-98.6 Mya). In this case, the estimated coalescence date for the *S. paradoxus* sequences (north vs. south) was 0.60 (95% CI of 0.33-0.94) Mya.

Discussion

There may be a number of potential caveats when considering the relationships among some taxa or the molecular dating: (1) there is a potential that some published mitogenomic sequences may have unwittingly incorporated numts (Lopez et al. 1994); (2) the analyses may be affected by different evolutionary rates across branches; (3) long branch attraction may potentially affect relationships; (4) various factors may lead to mito-nuclear incongruence, notably hybridization and sex-biased gene flow (Li et al. 2016). Given those potential caveats, it is noteworthy that using mammalian mitogenomes, the solenodons are estimated to have diverged from other mammals ca. 78.2 Mya (95% CI of 62.1-98.6 Mya). This is remarkably close to the previous point

estimate of 76 Mya for divergence initially estimated using primarily nuclear gene sequences (Roca *et al.* 2004), and to a more recent estimate based on nuclear sequences of 77.3 Mya (95% CI of 70.7-85.8 Mya) (Meredith et al. 2011).

This time estimate would be consistent with a vicariant origin for solenodons, since the proto-Antilles may have been connected to the North American continent some 70-80 Mya (Hedges 1996; Iturralde-Vinent and MacPhee 1999; Acton et al. 2000). However, the end of the Cretaceous 65 Mya is marked by the impact of an asteroid at nearby Chicxulub in the Yucatan (Iturralde-Vinent and MacPhee 1999). Small fossorial mammals may be more likely to survive some catastrophic events (Andersen and MacMahon 1985), and it is unlikely that islands of the proto-Antilles were completely inundated by the tsunami caused by the shallow-sea impact (Bryant 2014). Nonetheless, any life in the proto-Antilles would have faced catastrophic effects, especially since the islands would have been even closer to the impact site than they are today (Hedges 1996). Additionally, although some geologists have stated that the mountains of Hispaniola have remained above sea level continuously since the Mesozoic (Donnelly 1992), others have argued that every Caribbean island has been submerged at some point during the Cenozoic (Iturralde-Vinent and MacPhee 1999), which if true would rule out a completely vicariant origin for the solenodon lineage.

A Mesozoic divergence date versus other mammals does not necessarily imply a vicariant origin, as the solenodon lineage may have diverged more recently from mainland mammals that are now extinct. Some North American extinct taxa thought to be related to *Solenodon*, such as *Centetodon* or *Apternodus* (Asher et al. 2002; Lopatin 2003), did not appear to be close to *Solenodon* when morphological datasets were

analyzed while constrained to a scaffold tree of known molecular relationships (Asher et al. 2002; Roca et al. 2004). The recently extinct Nesophontidae (Whidden and Asher 2001; Rzebik-Kowalska and Woloszyn 2012; Orihuela 2014), a family of 11 species of insectivores that also lived only in the islands of the West Indies but went extinct after European settlement, also did not appear close to solenodons in that analysis. One issue with a dispersal of solenodons into the Antilles is that they are present only in Cuba and Hispaniola, with their origin in the two islands attributed to vicariance, since the divergence time between the two species (Roca et al. 2004) occurred when the two islands separated geologically (Iturralde-Vinent and MacPhee 1999). Since solenodons are also not found in other Caribbean islands, they are unlikely to be natural dispersers. *Nesophontes* appears to have more regularly dispersed across islands, as species were present in Puerto Rico and the Cayman Islands for which dispersal is the only plausible explanation (Whidden and Asher 2001). Recently, an analysis of endocranial morphology has suggested similarities between *Solenodon* and *Nesophontes* (Orihuela 2014). If a sister relationship is upheld between *Solenodon* and *Nesophontes*, one can hypothesize that the ancestral form was more likely to resemble *Nesophontes*, for which there is strong evidence for dispersal, and that the morphology of non-dispersing *Solenodon* may be derived subsequent to arrival.

Based on an examination of 65 southern and 128 northern specimens of *S. paradoxus*, Ottenwalder has proposed placing populations in the Peninsula de Barahona and Sierra de Bahoruco of the southwest Dominican Republic in a distinct subspecies *S. p. woodi*, characterized by its small body size, and into which he also placed the solenodons from Massif de la Hotte in Haiti (Ottenwalder 2001). Solenodons living north

of the Cul de Sac Plain/Neiba Valley (see Figure 1) were placed into the subspecies *S. p. paradoxus*. These differences were likely due to repeated separation of northern and southern Hispaniola by a marine canal into north and south paleo-islands during the Pliocene and Pleistocene (Ottenwalder 2001). Our results are consistent with the proposed subspecies division, since solenodons from the northern and southern Dominican Republic (Figure 1) separate in a haplotype network (Figure 3C). The southern Hispaniola solenodons appear to have much less genetic diversity than those in the north (Figure 3C), which may be related to the sixfold larger size of the northern paleo-island over the southern paleo-island (Figure 1) (Ottenwalder 2001). The estimated F_{ST} of 0.72 between northern and southern populations suggests a high degree of differentiation between the two geographic groups. This division into two subspecies would have needed confirmed by additional sampling and by the use of nuclear genetic markers. As this manuscript was in the peer-review process, a broader survey of solenodons across Hispaniola was reported (Turvey et al. 2016), based on 534 bp of the control region and 411 bp of cytochrome b in 34 samples of *Solenodon paradoxus*. They report the species to be genetically divided into northern, south-eastern and south-western populations (the last being confined to the Massif de la Hotte in Haiti) (Turvey et al. 2016). They also report morphometric support for this subdivision using 110 specimens (Turvey et al. 2016). Our results here thus confirm their finding that the northern and south-eastern populations are distinctive, and that the south-eastern population has very limited mitochondrial DNA diversity, and provide further support for the conservation management of Hispaniolan solenodons as distinct regional subspecies or evolutionarily significant units.

Acknowledgements

We thank Nicolas Corona and Yimell Corona for assistance in collecting samples.

Declaration of Interest Statement

The authors have declared that there is no conflict of interest.

Funding

SJO, AK, and PD were supported by the Russian Ministry of Science Mega-grant no.11.G34.31.0068.

References

Acton GD, Galbrun B, King JW. 2000. Paleolatitude of the Caribbean Plate since the Late Cretaceous. In: Leckie RM, Sigurdsson H, Acton GD, Draper G, editors. Proc. ODP, Sci. Results. College Station (TX): Ocean Drilling Program. p. 149-173.

Allard MW, Baker SD, Emerson GL, Ottenwalder JA, Kilpatrick CW. 2001. Characterization of the mitochondrial control region of *Solenodon paradoxus*

405 from Hispaniola and the implications for biogeography, systematics, and
406 conservation management. In: Woods CA, Sergile FE, editors. Biogeography of
407 the West Indies: Patterns and Perspectives. 2nd ed. Boca Raton (FL): CRC Press.
408 p. 331-334.
409 Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment
410 search tool. J Mol Biol. 215:403-10.
411 Andersen DC, MacMahon JA. 1985. The effects of catastrophic ecosystem disturbance:
412 the residual mammals at Mount St. Helens. J Mammal. 66:581-589.
413 Asher RJ, McKenna MC, Emry RJ, Tabrum AR, Kron DG. 2002. Morphology and
414 relationships of *Apternodus* and other extinct, zalambdodont placental mammals.
415 Bul Amer Mus Nat Hist. 217:1-117.
416 Bandelt HJ, Yao YG, Bravi CM, Salas A, Kivisild T. 2009. Median network analysis of
417 defectively sequenced entire mitochondrial genomes from early and contemporary
418 disease studies. J Hum Genet. 54(3):174-81.
419 Borroto-Paez R. 2009. Invasive mammals in Cuba: An overview. Biol Inv. 11(10):2279-
420 2290.
421 Bryant E. 2014. Tsunami: The Underrated Hazard. 3rd ed. London: Springer.
422 Butler PM. 1988. Phylogeny of the insectivores. In: Benton MJ, editor. The Phylogeny
423 and Classification of the Tetrapods. Oxford: Oxford University Press. p. 117-141.
424 Choi Y, Chan AP. 2015. PROVEAN web server: a tool to predict the functional effect of
425 amino acid substitutions and indels. Bioinformatics. 31(16):2745-7.
426 Darriba D, Taboada GL, Doallo R, Posada D. 2012. jModelTest 2: more models, new
427 heuristics and parallel computing. Nat Methods. 9:772.

428 Donnelly TW. 1992. Geological setting and tectonic history of Mesoamerica. In:
429 Quintero D, Aiello A, editors. Insects of Panama and Mesoamerica: selected
430 studies. Oxford: Oxford University Press. p. 1-13.
431 Drummond AJ, Suchard MA, Xie D, Rambaut A. 2012. Bayesian phylogenetics with
432 BEAUti and the BEAST 1.7. Mol Biol Evol. 29:1969-1973.
433 Eisenberg JF, Gould E. 1966. The behavior of *Solenodon paradoxus* in captivity with
434 comments on the behavior of other Insectivora. Zoologica. 51:49-58.
435 Eisenberg JF, Gozalez Gotera N. 1985. Observations on the natural history of *Solenodon*
436 *cubanus*. Acta Zoolog Fennica. 173:275-277.
437 Emerson GL, Kilpatrick CW, McNiff BE, Ottenwalder J, Allard MW. 1999.
438 Phylogenetic relationships of the order Insectivora based on complete 12S rRNA
439 sequences from mitochondria. Cladistics. 15:221-230.
440 Excoffier L, Lischer HE. 2010. Arlequin suite ver 3.5: a new series of programs to
441 perform population genetics analyses under Linux and Windows. Mol Ecol
442 Resour. 10:564-7.
443 Folinsbee KE. 2013. Evolution of venom across extant and extinct eulipotyphlans.
444 Comptes Rendus Palevol. 12:531-542.
445 Gernhard T. 2008. New analytic results for speciation times in neutral models. Bull Math
446 Biol. 70:1082-1097.
447 Goios A, Pereira L, Bogue M, Macaulay V, Amorim A. 2007. mtDNA phylogeny and
448 evolution of laboratory mouse strains. Genome Res. 17:293-298.

449 Hahn C, Bachmann L, Chevreux B. 2013. Reconstructing mitochondrial genomes
450 directly from genomic next-generation sequencing reads–a baiting and iterative
451 mapping approach. Nucleic Acids Res. 41:e129.
452 Hedges SB. 1996. Historical biogeography of West Indian vertebrates. Annu Rev Ecol
453 Syst. 27:163-196.
454 Iturralde-Vinent MA, MacPhee RDE. 1999. Paleogeography of the Caribbean region:
455 implications for Cenozoic biogeography. Bull Amer Mus Nat Hist. 238:1-238.
456 IUCN. 2013. The International Union for the Conservation of Nature (IUCN) Red List of
457 Threatened Species, 2013.2. http://www.redlist.org.
458 Katoh K, Standley DM. 2013. MAFFT multiple sequence alignment software version 7:
459 improvements in performance and usability. Mol Biol Evol. 30:772-80.
460 Krzywinski M, Schein J, Birol I, Connors J, Gascoyne R, Horsman D, Jones SJ, Marra
461 MA. 2009. Circos: an information aesthetic for comparative genomics. Genome
462 Res. 19:1639-45.
463 Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. Nat Meth.
464 9:357-9.
465 Li G, Davis BW, Eizirik E, Murphy WJ. 2016. Phylogenomic evidence for ancient
466 hybridization in the genomes of living cats (Felidae). Genome Res. 26:1-11.
467 Li H. 2011. A statistical framework for SNP calling, mutation discovery, association
468 mapping and population genetical parameter estimation from sequencing data.
469 Bioinformatics. 27:2987-2993.

Lopatin AV. 2003. A zalambdodont insectivore of the family Apternodontidae (Insectivora, Mammalia) from the Middle Eocene of Mongolia. Paleontol J. 37:187-195.

Lopez JV, Yuhki N, Masuda R, Modi W, O'Brien SJ. 1994. Numt, a recent transfer and tandem amplification of mitochondrial DNA to the nuclear genome of the domestic cat. J Mol Evol. 39(2):174-90.

Loytynoja A, Goldman N. 2010. webPRANK: a phylogeny-aware multiple sequence aligner with interactive alignment browser. BMC Bioinformatics. 11:579.

MacPhee RDE, Flemming C, Lunde DP. 1999. "Last occurrence" of the Antillean insectivoran *Nesophontes*: new radiometric dates and their interpretation. Amer Mus Novit. 3261:1-20.

MacPhee RDE, Novacek MJ. 1993. Definitions and relationships of Lipotyphla. In: Szalay FS, Novacek MJ, McKenna MC, editors. Mammal Phylogeny: Placentals. New York: Springer-Verlag. p. 13-31.

McDowell SB, Jr. 1958. The Greater Antillean insectivores. Bull Amer Mus Nat Hist. 115:113-214.

McKenna MC, Bell SK. 1997. Classification of Mammals above the Species Level. New York: Columbia University Press.

McWilliam H, Li WZ, Uludag M, Squizzato S, Park YM, Buso N, Cowley AP, Lopez R. 2013. Analysis Tool Web Services from the EMBL-EBI. Nucleic Acids Res. 41:W597-W600.

491 Meredith RW, Janecka JE, Gatesy J, Ryder OA, Fisher CA, Teeling EC, Goodbla A,
492 Eizirik E, Simao TL, Stadler T, et al. 2011. Impacts of the Cretaceous terrestrial
493 revolution and KPg extinction on mammal diversification. Science. 334:521-524.
494 Miller MA, Schwartz T, Pickett BE, He S, Klem EB, Scheuermann RH, Passarotti M,
495 Kaufman S, O'Leary MA. 2015. A RESTful API for access to phylogenetic tools
496 via the CIPRES Science Gateway. Evol Bioinform Online. 11:43-48.
497 Mouchaty SK, Gullberg A, Janke A, Arnason U. 2000. Phylogenetic position of the
498 tenrecs (Mammalia: Tenrecidae) of Madagascar based on analysis of the complete
499 mitochondrial genome sequence of *Echinops telfairi*. Zool Scr. 29:307-317.
500 Orihuela J. 2014. Endocranial morphology of the extinct Antillean shrew *Nesophontes*
501 (Lipotyphla: Nesophontidae) from natural and digital endocasts of Cuban taxa.
502 Palaeontologia Electronica. 17:22A.
503 Ottenwalder J. 2001. Systematics and biogeography of the West Indian genus *Solenodon.*
504 In: Woods CA, Sergile FE, editors. Biogeography of the West Indies: Patterns and
505 Perspectives. 2nd ed. Boca Raton (FL): CRC Press. p. 253-329.
506 Ottenwalder JA. 1985. The distribution and habitat of *Solenodon* in the Dominican
507 Republic. [MS Thesis]. Gainesville (FL): University of Florida.
508 Roca AL, Bar-Gal GK, Eizirik E, Helgen KM, Maria R, Springer MS, O'Brien SJ,
509 Murphy WJ. 2004. Mesozoic origin for West Indian insectivores. Nature.
510 429:649-651.
511 Rzebik-Kowalska B, Woloszyn BW. 2012. New data on *Nesophontes* subfossil
512 populations from Cuba and Isla de la Juventud (Cuba). Neues Jahrbuch für
513 Geologie und Paläontologie. 263:155-166.

514 Schattner P, Brooks AN, Lowe TM. 2005. The tRNAscan-SE, snoscan and snoGPS web
515 servers for the detection of tRNAs and snoRNAs. Nucleic Acids Res. 33:W686-
516 W689.
517 Slater GS, Birney E. 2005. Automated generation of heuristics for biological sequence
518 comparison. BMC Bioinformatics. 6:31.
519 Soares P, Ermini L, Thomson N, Mormina M, Rito T, Rohl A, Salas A, Oppenheimer S,
520 Macaulay V, Richards MB. 2009. Correcting for purifying selection: an improved
521 human mitochondrial molecular clock. Am J Hum Genet. 84:740-759.
522 Stamatakis A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis
523 of large phylogenies. Bioinformatics. 30:1312-1313.
524 Stanhope MJ, Waddell VG, Madsen O, de Jong W, Hedges SB, Cleven GC, Kao D,
525 Springer MS. 1998. Molecular evidence for multiple origins of Insectivora and for
526 a new order of endemic African insectivore mammals. Proc Natl Acad Sci U S A.
527 95:9967-9972.
528 Starostina E, Tamazian G, Dobrynin P, O'Brien SJ, Komissarov A. 2015. Cookiecutter: a
529 tool for kmer-based read filtering and extraction. bioRxiv DOI:
530 http://dx.doi.org/10.1101/024679.
531 Turvey ST. 2010. Evolution of non-homologous venom delivery systems in West Indian
532 insectivores. J Vert Paleontol. 30:1294-1299.
533 Turvey ST, Fernández-Secades C, Nuñez-Miño JM, Hart T, Martinez P, Brocca JL,
534 Young RP. 2014. Is local ecological knowledge a useful conservation tool for
535 small mammals in a Caribbean multicultural landscape? Biol Conserv. 169:189-
536 197.

Turvey ST, Meredith HMR, Scofield RP. 2008. Continued survival of Hispaniolan solenodon *Solenodon paradoxus* in Haiti. Oryx. 42:611-614.

Turvey ST, Peters S, Brace S, Young RP, Crumpton N, Hansford J, Nunez-Mino JM, King G, Tsalikidis K, Ottenwalder JA, et al. 2016. Independent evolutionary histories in allopatric populations of a threatened Caribbean land mammal. Diversity Distrib. DOI: http://dx.doi.org/10.1111/ddi.12420

Waddell PJ, Shelley S. 2003. Evaluating placental inter-ordinal phylogenies with novel sequences including RAG1, gamma-fibrinogen, ND6, and mt-tRNA, plus MCMC-driven nucleotide, amino acid, and codon models. Mol Phylogenet Evol. 28:197-224.

Wheeler DL, Church DM, Federhen S, Lash AE, Madden TL, Pontius JU, Schuler GD, Schriml LM, Sequeira E, Tatusova TA et al. 2003. Database resources of the National Center for Biotechnology. Nucleic Acids Res. 31:28-33.

Whidden HP, Asher RJ. 2001. The origin of the Greater Antillean insectivorans. In: Woods CA, Sergile FE, editors. Biogeography of the West Indies: Patterns and Perspectives. 2nd ed. Boca Raton (FL): CRC Press. p. 237-252.

Wible JR. 2008. On the cranial osteology of the Hispaniolan Solenodon, *Solenodon paradoxus* Brandt, 1833 (Mammalia, Lipotyphla, Solenodontidae). Ann Carnegie Mus. 77:321-385.

Woods CA, Sergile FE. 2001. Biogeography of the West Indies: Patterns and Perspectives. Boca Raton (FL): CRC Press.

560 **Figure Legends**

561

562 **Figure 1.** Elevation map of Hispaniola showing sampling localities. Hispaniolan
563 solenodon mtDNA sequences were from samples from the localities indicated, in both the
564 southern and northern Dominican Republic (DR). From the southern DR, in the Province
565 of Pedernales, La Cañada del Verraco and two collection points in El Manguito were
566 sampled for the current study. Novel mitogenomic sequences were also obtained from a
567 sample previously collected from the Cordillea Septentrional in the Province of Espaillat
568 in the northern DR (Roca et al. 2004). Also included in some analyses were previously
569 generated mtDNA sequences (Emerson et al. 1999; Allard et al. 2001) from two
570 specimens from Loma de la Jagua, Cabrera Promontory, Province of Maria Trinidad
571 Sanchez in the northern DR; a third mtDNA sequence from the same study was from a
572 specimen of unknown provenance. The key for elevation above sea level is shown; the
573 darkest shaded regions have the lowest elevation. The dashed line shows the position of
574 the Cul de Sac Plain and Neiba Valley; this region was periodically inundated by a
575 marine canal that separated Hispaniola into north and south paleo-islands during the
576 Pliocene and Pleistocene (Ottenwalder 2001). The original map courtesy of NASA is in
577 the public domain.

578

579 **Figure 2.** Structure and variation in the mitogenomes of *Solenodon*. The thicker gray
580 circle shows the location of genes and the control region of a consensus of the 5
581 mitogenomes of individuals from the south, with the positions of the rRNA genes (red),
582 protein coding genes (blue), tRNA genes (orange) and control region (green) indicated.

583 Genome coordinates are in kilobases. The inner gray semi-circles indicate mitogenomic
584 variation among solenodons from the southern Dominican Republic (subspecies *S.*
585 *paradoxus woodi*) when compared to the consensus. Spa-K and -L are from La Cañada
586 del Verraco, while Spa-M, -N and -O are from nearby El Manguito. Very little variation
587 is seen among these solenodons from southern Hispaniola. By contrast, the mitogenome
588 of a solenodon from northern Hispaniola (subspecies *S. paradoxus paradoxus*),
589 designated Spa-1, shows much higher variability when compared to the southern
590 consensus sequence. At the top of the figure, fragments of mtDNA previously reported
591 (Roca et al. 2004) for the Cuban solenodon (*S. cubanus*) show even greater variation
592 relative to the reference mitogenome; insertions are shown as cut-ins on the tracks.
593
594 **Figure 3.** Relationship among Hispaniolan solenodons (*S. paradoxus*) from different
595 geographic regions. (A) Phylogeny generated using the full mitogenomic sequences,
596 including the control region (16,457 bp). Individuals are identified by sample number for
597 solenodons captured in the north (N) and south (S) of Hispaniola. The divergence date of
598 nodes was estimated using the rate observed in humans. Coalescence dates (in thousands
599 of years, with 95% CI) are indicated at the nodes with a chronogram below the tree. (B)
600 In order to root the tree, novel mtDNA sequences from *S. paradoxus* were aligned to a
601 previously reported 2,495 bp sequence of the Cuban solenodon (*S. cubanus*), which was
602 used as an outgroup (GenBank accession AY530083) (Roca et al. 2004). The root
603 separated the single northern Hispaniolan solenodon (Spa-1) from five southern
604 Hispaniolan solenodons. (C) To examine whether mitochondrial sequences of
605 Hispaniolan solenodons are different in the north and south of Hispaniola, the novel

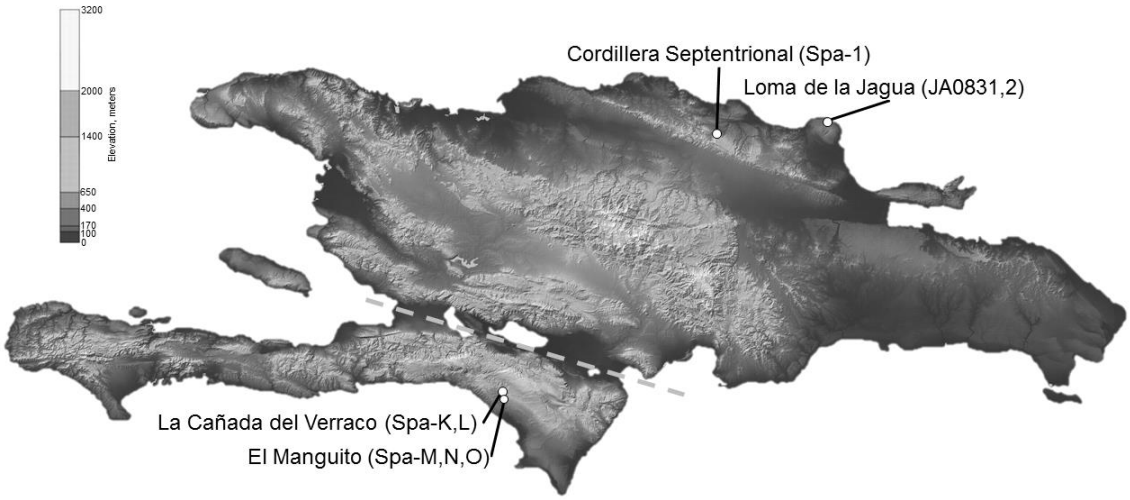
606 sequences were combined in a 389 bp alignment with previously generated control region
607 sequences from two specimens from the northern Dominican Republic, and one of
608 unknown provenance. The network shows each difference as a hatch mark (all were
609 substitutions, there were no indels). The control region sequences from the 5 southern
610 specimens were identical and at one end of the network, separated from the sequences of
611 northern Hispaniolan solenodons. The separation of northern and southern haplotypes is
612 consistent with a previous suggestion, based on morphological measurements, that *S.*
613 *paradoxus* be subdivided into a northern (*S. p. paradoxus*) and a southern (*S. p. woodi*)
614 subspecies (Ottenwalder 2001).
615

This is an Accepted Manuscript of an article published by Taylor & Francis Group in Mitochondrial DNA Part A on 15/03/2016, available online: http://dx.doi.org/10.3109/24701394.2016.1167891.

616 [[Figure 1]]

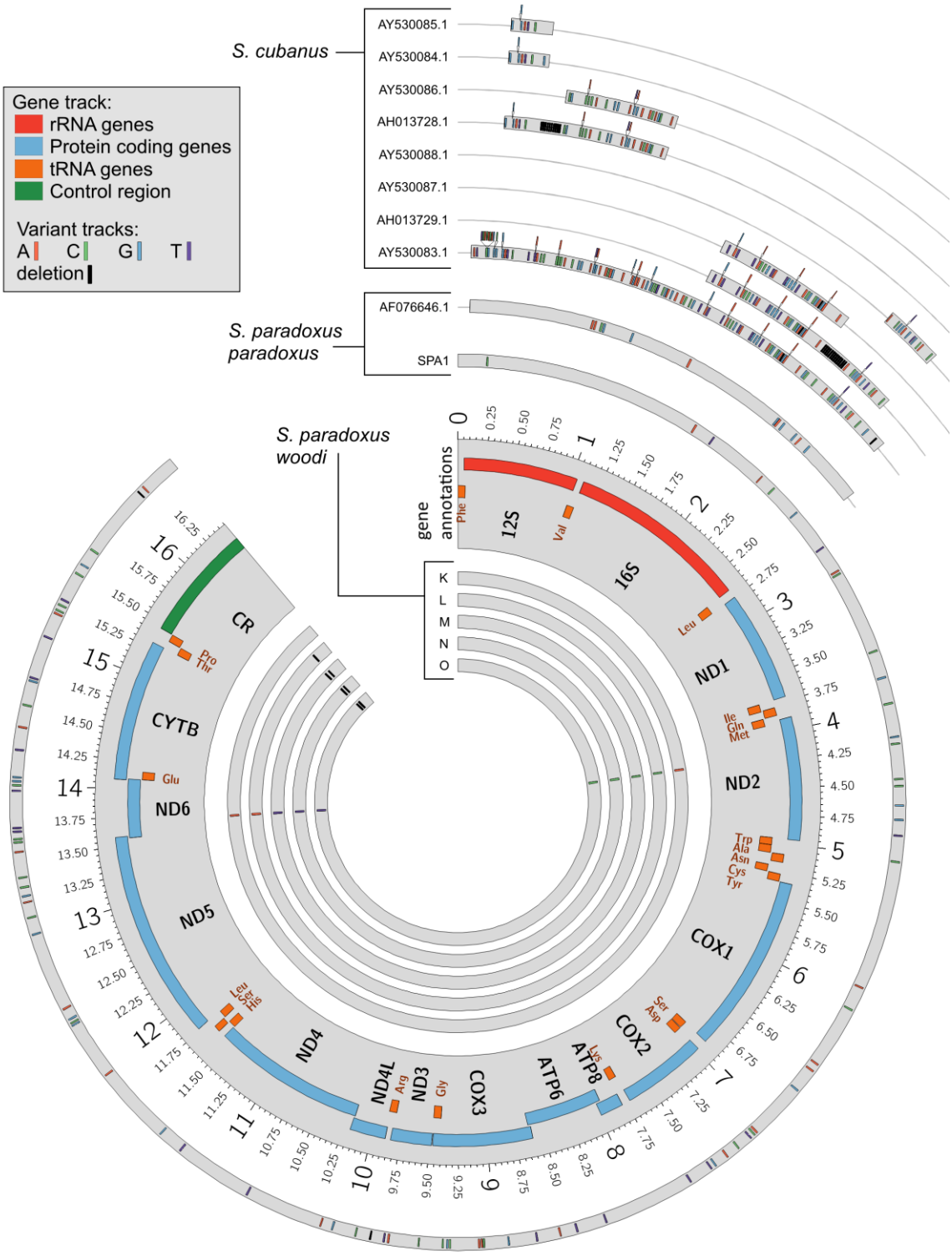
617

618



This is an Accepted Manuscript of an article published by Taylor & Francis Group in Mitochondrial DNA Part A on 15/03/2016, available online: http://dx.doi.org/10.3109/24701394.2016.1167891.

619 [[Figure 2]]



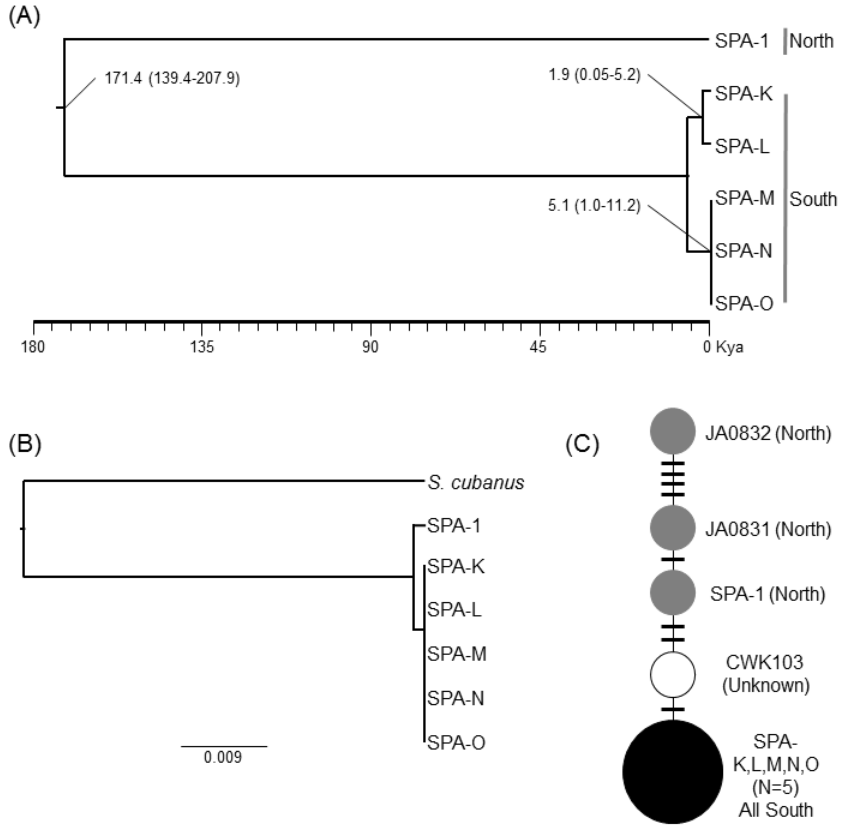
620

This is an Accepted Manuscript of an article published by Taylor & Francis Group in Mitochondrial DNA Part A on 15/03/2016, available online: http://dx.doi.org/10.3109/24701394.2016.1167891.

621 [[Figure 3]]

622

623



This is an Accepted Manuscript of an article published by Taylor & Francis Group in Mitochondrial DNA Part A on 15/03/2016, available online: http://dx.doi.org/10.3109/24701394.2016.1167891.

Table 1. Provenance of Solenodon paradoxus samples

ID	Region	Province	Locality	Sex	Collected	Weight (g)	Coordinates	Sequences
SPA-K	South	Pedernales	La Cañada del Verraco	M	Wild	579	N 18º 09' 9.64" W 71º 43' 12.0"	This study
SPA-L	South	Pedernales	La Cañada del Verraco	M	Wild	1020	N 18º 09' 9.64" W 71º 43' 12.0"	This study
SPA-M	South	Pedernales	El Manguito - 1	M	Wild	1270	N 18º 06' 36.6" W 71º 43' 3.58"	This study
SPA-N	South	Pedernales	El Manguito - 1	F	Wild	1420	N 18º 06' 36.6" W 71º 43' 3.58"	This study
SPA-O	South	Pedernales	El Manguito - 2	F	Wild	1120	N 18º 07' 6.5" W 71º 43' 14.7"	This study
SPA-1	North	Espaillat	Cordillera Septentrional	M	Zoo*	-	- -	This study
JA0832	North	Maria Trinidad Sanchez	Loma de la Jagua	-	-	-	- -	Allard et al. 2001
JA0831	North	Maria Trinidad Sanchez	Loma de la Jagua	-	-	-	- -	Allard et al. 2001
CWK103	Unknown	-	-	-	-	-	- -	Allard et al. 2001

*Collected at ZOODOM for a previous study not involving mitogenomics, Roca et al. 2004

This is an Accepted Manuscript of an article published by Taylor & Francis Group in Mitochondrial DNA Part A on 15/03/2016, available online: http://dx.doi.org/10.3109/24701394.2016.1167891.

Mitogenomic sequences support a north-south subspecies subdivision within *Solenodon paradoxus*

Supplementary Material

Authors: Adam L. Brandt[1,2+], Kirill Grigorev[4+], Yashira M. Afanador-Hernández[4], Liz A. Paulino[5], William J. Murphy[6], Adrell Núñez[7], Aleksey Komissarov[8], Jessica R. Brandt[1], Pavel Dobrynin[8], J. David Hernández-Martich[9], Roberto María[7], Stephen J. O'Brien[8,10], Luis E. Rodríguez[5], Juan C. Martínez-Cruzado[4], Taras K. Oleksyk[4*] and Alfred L. Roca[1,2,3*]

[+]Equal contributors
*Corresponding authors: roca@illinois.edu; taras.oleksyk@upr.edu

Affiliations: [1]Department of Animal Sciences, [2]Illinois Natural History Survey and [3]Carl R. Woese Institute for Genomic Biology, University of Illinois at Urbana-Champaign, Urbana, IL 61801; [4]Department of Biology, University of Puerto Rico at Mayagüez, Mayagüez, Puerto Rico 00681; [5]Instituto Tecnológico de Santo Domingo (INTEC), Santo Domingo, Dominican Republic; [6]Department of Veterinary Integrative Biosciences, Texas A&M University, College Station, TX 77843; [7]Department of Conservation and Science, Parque Zoológico Nacional (ZOODOM), Santo Domingo, Dominican Republic; [8]Theodosius Dobzhansky Center for Genome Bioinformatics, St. Petersburg State University, 199004 St. Petersburg, Russia; [9]Instituto de Investigaciones Botanicas y Zoologicas, Universidad Autónoma de Santo Domingo, Santo Domingo, Dominican Republic; [10]Oceanographic Center, Nova Southeastern University, Fort Lauderdale, Florida 33004

Figure S1. Relationship of solenodon mitochondrial genomes to those of other mammals. The maximum likelihood tree of laurasiatherian mammals showing the position of solenodons was generated in RAxML version 8.2.4 (Stamatakis 2014), using the GTR+I+G substitution model and 1000 rapid bootstrap replicates. Full mitochondrial genome sequences (excluding control region and regions of poor alignment) from the taxa on the tree were used to generate the phylogeny. Letters indicate the nodes for which fossil calibration dates and inferred molecular dates are described in Table S3.

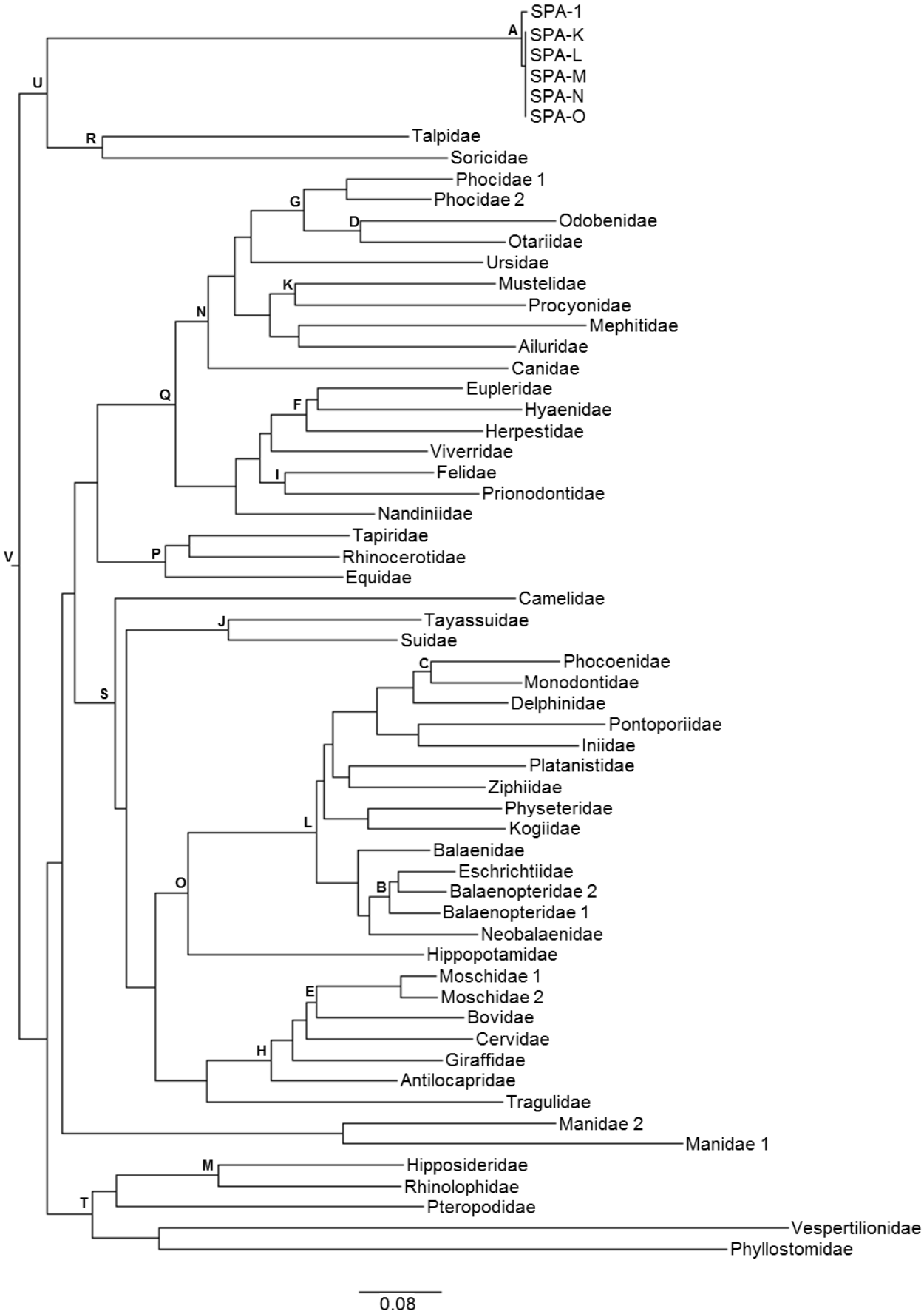
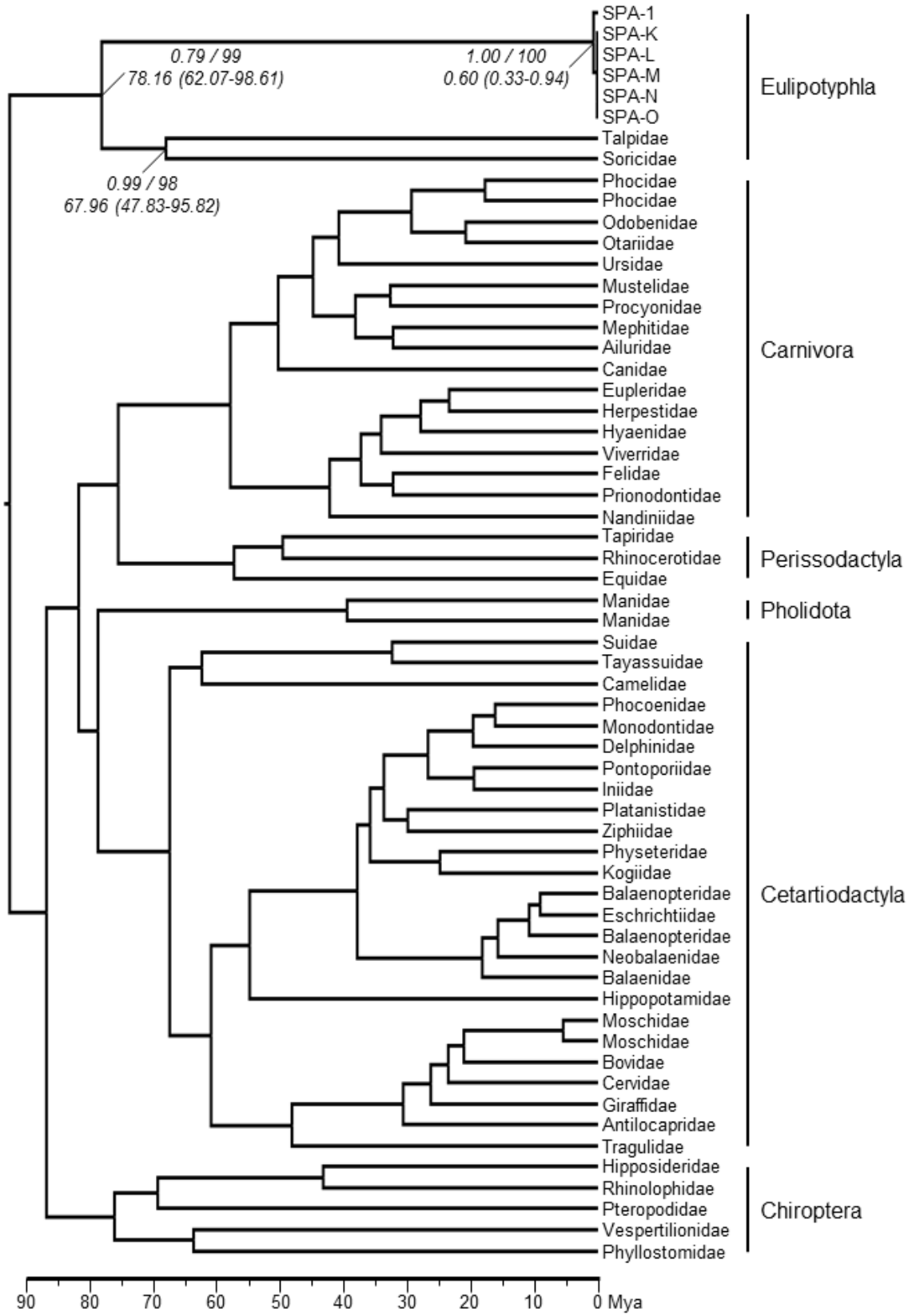


Figure S2. Phylogeny of laurasiatherian mammals showing the position of *Solenodon paradoxus*. The tree was generated using the full mitogenome sequences except that regions of poor alignment were excluded, as was the control region, which is unreliable for the deep time scales involved (Ingman et al. 2000). The relationships among taxa were not constrained to those established using the fossil record or nuclear genetic relationships, since in some circumstances mitogenomic relationships may be incongruent with the true species tree (Petit and Excoffier 2009). For molecular dating, previously well-established fossil calibrations (Meredith et al. 2011) were incorporated into an analysis conducted using the software BEAST v1.8.2 software (Drummond et al. 2012). At eulipotyphlan nodes, the Bayesian posterior probability and maximum likelihood bootstrap support are listed above, while below are listed dates of divergence with 95% credibility intervals in parentheses. The time at which the solenodon mitogenome diverged from that of other mammals was estimated as 78.2 Mya (95% CI: 62.1-98.6 Mya). The point estimate for the divergence time between solenodons and other mammals is similar to previous estimates ranging from 73-76 Mya, based upon analysis of datasets largely comprised of nuclear sequences (Roca et al. 2004, Meredith et al. 2011).

This is an Accepted Manuscript of an article published by Taylor & Francis Group in Mitochondrial DNA Part A on 15/03/2016, available online: http://dx.doi.org/10.3109/24701394.2016.1167891.



This is an Accepted Manuscript of an article published by Taylor & Francis Group in Mitochondrial DNA Part A on 15/03/2016, available online: http://dx.doi.org/10.3109/24701394.2016.1167891.

Table S1. Radical mutations in mitochondrial protein coding genes of *S. paradoxus* in comparison to those of reference species. Mutations are shown in this format: position, amino acid in the reference mitogenome, amino acid in *S. paradoxus*. Scores were calculated with PROVEAN using a stringent threshold for listing the radical mutations of -4.1. Entries are vertically aligned with regard to preceding insertion and deletion events.

Gene	Reference species							
	Talpa europaea		***Sorex araneus***		***Felis catus***		***Homo sapiens***	
ATP6							33,T,H	-4.310
	125,L,T	-4.525	125,L,T	-4.478	125,L,T	-4.537	125,L,T	-4.453
ATP8					24,F,L	-4.253		
	27,K,A	-5.702	27,K,A	-5.677	27,K,A	-5.755	27,K,A	-5.993
	34,P,T	-4.805	34,P,T	-4.792	34,P,T	-4.690	34,H,T	-4.171
					39,P,L	-4.293		
	62,L,H	-5.634	62,L,H	-5.016	62,L,H	-4.682		
COX1	–	–	–	–	–	–	–	–
COX2	39,H,G	-7.301			39,H,G	-7.099		
							151,P,T	-5.157
COX3	–	–	–	–	–	–	–	–
CYTB	–	–	–	–	–	–	–	–
ND1	252,P,S	-4.615	252,P,S	-5.249	252,P,S	-5.008	250,P,S	-5.384
	308,P,H	-6.980	308,P,H	-7.014	308,P,H	-6.881	306,P,H	-7.051
ND2			20,G,S	-4.458	25,G,S	-4.367	22,G,S	-5.039
	46,M,T	-4.480	41,M,T	-4.937			43,M,T	-4.721
			96,P,I	-5.200	101,P,I	-5.685		
			147,Y,M	-5.460	152,Y,M	-6.104	149,Y,M	-5.543
							153,P,N	-5.191
			202,P,Q	-6.613	207,P,Q	-6.447	204,P,Q	-6.187
	243,W,G	-10.996	238,W,G	-10.116	243,W,G	-10.578	240,W,G	-10.553
	244,N,S	-4.510	239,N,S	-4.366	244,N,S	-4.544	241,N,S	-4.500
	254,L,T	-4.230						
							295,N,S	-4.223
	313,M,T	-4.258	308,M,T	-4.449	313,M,T	-4.215		
	314,F,L	-4.986	309,F,L	-5.164	314,F,L	-5.218		
ND3	27,L,P	-4.797	27,L,P	-5.226	27,L,P	-4.965	27,L,P	-5.285
ND4							4,H,M	-4.380
							5,S,Y	-4.341
							185,G,S	-4.427
ND4L	–	–	–	–	–	–	–	–
ND5	62,S,Y	-4.845	62,S,Y	-4.857	62,S,Y	-4.745	60,S,Y	-4.623
	154,G,S	-4.377	154,G,S	-4.401	154,G,S	-4.377		
	349,N,A	-6.467	349,N,A	-6.734	349,N,A	-6.467	347,N,A	-6.667
	431,G,K	-4.196	431,G,K	-4.371			429,G,K	-4.393
	435,F,S	-5.564			435,F,S	-5.562	433,F,S	-5.628
	446,P,K	-5.870	446,P,K	-5.925			444,P,K	-5.883
	450,N,K	-4.681	450,N,K	-4.558	450,N,K	-4.681	448,N,K	-4.691
	469,N,Y	-4.690	469,N,Y	-4.226	469,N,Y	-4.560	467,N,Y	-4.369
	510,K,M	-4.121	510,K,M	-4.451	510,K,M	-4.244	508,K,M	-4.423
			514,P,K	-4.402	514,P,K	-4.325	512,P,K	-4.544
	578,Q,N	-4.282	578,Q,N	-4.335	578,Q,N	-4.282	576,Q,N	-4.294
ND6	106,Y,C	-4.909	106,Y,C	-4.378	106,Y,C	-4.945	106,W,C	-5.745
	138,E,G	-4.758			138,E,G	-4.775		
							162,F,V	-5.011

Table S2. Taxa used for comparison of *S. paradoxus* to other mammals.

Order	Family	Taxon	GenBank accession code
Eulipotyphla	Solenodontidae	SPA-1	KU697358
Eulipotyphla	Solenodontidae	SPA-K	KU697359
Eulipotyphla	Solenodontidae	SPA-L	KU697360
Eulipotyphla	Solenodontidae	SPA-M	KU697361
Eulipotyphla	Solenodontidae	SPA-N	KU697362
Eulipotyphla	Solenodontidae	SPA-O	KU697363
Eulipotyphla	Soricidae	European shrew	NC_027963
Eulipotyphla	Talpidae	European mole	NC_002391
Cetartiodactyla	Antilocapridae	Pronghorn	NC_020679
Cetartiodactyla	Balaenidae	Southern right whale	NC_006930
Cetartiodactyla	Balaenopteridae 1	Pygmy Bryde's whale	NC_007938
Cetartiodactyla	Balaenopteridae 2	Humpback whale	NC_006927
Cetartiodactyla	Bovidae	Domestic cattle	NC_006853
Cetartiodactyla	Camelidae	Bactrian camel	NC_009628
Cetartiodactyla	Cervidae	Reindeer	NC_007703
Cetartiodactyla	Delphinidae	Heaviside's dolphin	NC_020696
Cetartiodactyla	Eschrichtiidae	Grey whale	NC_005270
Cetartiodactyla	Giraffidae	Giraffe	NC_024820
Cetartiodactyla	Hippopotamidae	Pygmy hippopotamus	NC_020697
Cetartiodactyla	Iniidae	Amazon river dolphin	NC_005276
Cetartiodactyla	Kogiidae	Pygmy sperm whale	NC_005272
Cetartiodactyla	Monodontidae	Narwhal	NC_005279
Cetartiodactyla	Moschidae 1	Anhui musk deer	NC_020017
Cetartiodactyla	Moschidae 2	Siberian musk deer	NC_013753
Cetartiodactyla	Neobalaenidae	Pygmy right whale	NC_005269
Cetartiodactyla	Phocoenidae	Finless porpoise	NC_026456
Cetartiodactyla	Physeteridae	Sperm whale	NC_002503
Cetartiodactyla	Platanistidae	Indus River dolphin	NC_005275
Cetartiodactyla	Pontoporiidae	La Plata dolphin	NC_005277
Cetartiodactyla	Suidae	Domestic pig	NC_000845
Cetartiodactyla	Tayassuidae	Collared peccary	NC_012103
Cetartiodactyla	Tragulidae	Water chevrotain	NC_020714
Cetartiodactyla	Ziphiidae	Ginkgo-toothed beaked whale	NC_027593
Perissodactyla	Equidae	Horse	NC_001640
Perissodactyla	Rhinocerotidae	White rhinoceros	NC_001808
Perissodactyla	Tapiridae	Malayan tapir	NC_023838
Carnivora	Ailuridae	Red panda	NC_009691
Carnivora	Canidae	Domestic dog	NC_002008

Table S2. Continued.

Order	Family	Taxon	GenBank accession code
Carnivora	Eupleridae	Narrow-striped mongoose	NC_027828
Carnivora	Felidae	Domestic cat	NC_001700
Carnivora	Herpestidae	Indian mongoose	NC_006835
Carnivora	Hyaenidae	Spotted hyena	NC_020670
Carnivora	Mephitidae	Striped skunk	NC_020648
Carnivora	Mustelidae	American marten	NC_020642
Carnivora	Nandiniidae	African palm civet	NC_024567
Carnivora	Odobenidae	Atlantic walrus	NC_004029
Carnivora	Otariidae	California sea lion	NC_008416
Carnivora	Phocidae 1	Harp seal	NC_008429
Carnivora	Phocidae 2	Weddell seal	NC_008424
Carnivora	Prionodontidae	Spotted linsang	NC_024569
Carnivora	Procyonidae	Raccoon	NC_009126
Carnivora	Ursidae	Giant short-faced bear	NC_011116
Carnivora	Viverridae	Servaline genet	NC_024568
Chiroptera	Hipposideridae	Great roundleaf bat	NC_018540
Chiroptera	Phyllostomidae	Common vampire bat	NC_022423
Chiroptera	Pteropodidae	Large flying fox	NC_026542
Chiroptera	Rhinolophidae	Big-eared horseshoe Bat	NC_026460
Chiroptera	Vespertilionidae	Red bat	NC_016873
Pholidota	Manidae 1	Tree pangolin	NC_026780
Pholidota	Manidae 2	Malayan pangolin	NC_026781

This is an Accepted Manuscript of an article published by Taylor & Francis Group in Mitochondrial DNA Part A on 15/03/2016, available online: http://dx.doi.org/10.3109/24701394.2016.1167891.

Table S3. Laurasiatherian fossil calibration dates and inferred molecular dating.

Group	Node	Fossil*	Estimated	(95% CI)	Posterior probability	Bootstrap node support (%)
Base of *S. paradoxus*	A	-	0.60	(0.33-0.94)	1.00	100
Megaptera-Eschrichtius	B	7.3-23.03	10.74	(7.69-14.02)	1.00	100
Phocoenidae-Monodontidae	C	12.1-28.5	16.10	(11.52-20.40)	1.00	100
Otariidae-Odobenidae	D	15.97-34	20.84	(15.83-25.67)	1.00	100
Bovidae-Moschidae	E	18-34	21.05	(16.49-25.54)	1.00	31
Herpestidae-Eupleridae	F	15.97-34	23.43	(18.44-28.85)	1.00	100
Phocidae-(Otariidae+Odobenidae)	G	20.43-34	29.36	(24.93-34.06)	1.00	100
Giraffidae-Antilocapridae	H	17.8-34	30.55	(25.80-35.39)	1.00	100
Felidae-Prionodontidae	I	28.3-40.6	32.27	(27.29-37.10)	1.00	99
Suidae-Tayassuidae	J	15.97-37.3	32.37	(35.55-41.13)	1.00	100
Procyonidae-Mustelidae	K	27.6-40.6	32.61	(27.50-37.43)	1.00	96
Cetacea	L	33.8-48.8	37.80	(33.27-42.18)	1.00	100
Hipposideridae-Rhinolophidae	M	37.1-56	43.28	(34.18-52.18)	1.00	100
Canidae-Arctoidea	N	37.1-56	50.38	(45.25-55.75)	1.00	100
Whippomorpha	O	52.5-61.1	54.80	(50.85-58.73)	1.00	100
Base of Perissodactyla	P	55.5-61.1	57.33	(54.11-60.58)	1.00	100
Base of Carnivora	Q	37.1-65.8	57.78	(52.11-63.57)	1.00	100
Talpidae-Soricidae	R	-	67.96	(47.83-95.82)	0.99	98
Base of Cetartiodactyla	S	52.5-65.8	67.46	(52.67-72.57)	1.00	100
Base of Chiroptera	T	-	76.09	(66.57-85.48)	1.00	100
Base of Eulipotyphla	U	-	78.16	(62.07-98.61)	0.79	99
Base of Laurasiatheria	V	-	92.61	(80.06-119.48)	1.00	-

*Fossil calibration dates are as described previously (Meredith *et al.* 2011).

Supplementary References

Drummond AJ, Suchard MA, Xie D, Rambaut A. 2012. Bayesian phylogenetics with BEAUti and the BEAST 1.7. Mol Biol Evol. 29:1969-1973.

Ingman M, Kaessmann H, Paabo S, Gyllensten U. 2000. Mitochondrial genome variation and the origin of modern humans. Nature. 408:708-713.

Meredith RW, Janecka JE, Gatesy J, Ryder OA, Fisher CA, Teeling EC, Goodbla A, Eizirik E, Simao TL, Stadler T, et al. 2011. Impacts of the Cretaceous Terrestrial Revolution and KPg extinction on mammal diversification. Science 334:521-524.

Petit RJ, Excoffier L. 2009. Gene flow and species delimitation. Trends Ecol Evol. 4:386-393.

Roca AL, Bar-Gal GK, Eizirik E, Helgen KM, Maria R, Springer MS, O'Brien SJ, Murphy WJ. 2004. Mesozoic origin for West Indian insectivores. Nature 429:649-651.

Stamatakis A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics. 30:1312-1313.