

St. Norbert College

Digital Commons @ St. Norbert College

Faculty Creative and Scholarly Works

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

Follow this and additional works at: https://digitalcommons.snc.edu/faculty_staff_works

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

1 *Mitochondrial DNA*, Original Article

2

3

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

6

7

8 Authors: Adam L. Brandt^{1,2+}, Kirill Grigorev⁴⁺, Yashira M. Afanador-Hernández⁴, Liz A.
9 Paulino⁵, William J. Murphy⁶, Adrell Núñez⁷, Aleksey Komissarov⁸, Jessica R. Brandt¹,
10 Pavel Dobrynin⁸, J. David Hernández-Martich⁹, Roberto María⁷, Stephen J. O'Brien^{8,10},
11 Luis E. Rodríguez⁵, Juan C. Martínez-Cruzado⁴, Taras K. Oleksyk^{4*} and Alfred L.
12 Roca^{1,2,3*}

13

14 ⁺Equal contributors

15 *Corresponding authors: roca@illinois.edu; taras.oleksyk@upr.edu

16

17

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

30

31

32 Corresponding address:

33 Alfred L. Roca, Ph.D.

34 Department of Animal Sciences

35 University of Illinois at Urbana-Champaign

36 432 ASL, 1207 West Gregory Drive

37 Urbana, Illinois 61801

38 Email: roca@illinois.edu

39 Mobile phone: 240-422-2493

40 Office phone: 217-244-8853

41 Department fax: 217-333-8286

42 **Abstract**

43 Solenodons are insectivores found only in Hispaniola and Cuba, with a Mesozoic
44 divergence date versus extant mainland mammals. Solenodons are the oldest lineage of
45 living eutherian mammal for which a mitogenome sequence has not been reported. We
46 determined complete mitogenome sequences for six Hispaniolan solenodons (*Solenodon*
47 *paradoxus*) using next generation sequencing. The solenodon mitogenomes were 16,454-
48 16,457 bp long and carried the expected repertoire of genes. A mitogenomic phylogeny
49 confirmed the basal position of solenodons relative to shrews and moles, with solenodon
50 mitogenomes estimated to have diverged from those of other mammals ca. 78 Mya.
51 Control region sequences of solenodons from the northern (n=3) and southern (n=5)
52 Dominican Republic grouped separately in a network, with $F_{ST} = 0.72$ ($p=0.036$) between
53 north and south. This regional genetic divergence supports previous morphological and
54 genetic reports recognizing northern (*S. p. paradoxus*) and southern (*S. p. woodi*)
55 subspecies in need of separate conservation plans.

56

57

58 **Key words [3-6]:** conservation, Dominican Republic, evolutionarily significant units,

59 *Solenodon paradoxus*, subspecies, taxonomy

60 **Introduction**

61

62 Solenodons are fossorial insectivores that live only in the islands of Cuba and
63 Hispaniola (Ottewalder 1985). The two species of solenodon are among the few native
64 non-volant mammals to survive human colonization of the West Indies (MacPhee et al.
65 1999; Ottewalder 2001). The geographic range of the Cuban solenodon (*Solenodon*
66 *cubanus* Peters, 1861) is currently limited to eastern Cuba (Ottewalder 1985). The
67 Hispaniolan solenodon (*S. paradoxus* Brandt, 1833) lives only in the island of
68 Hispaniola, with surviving range primarily within the Dominican Republic (Ottewalder
69 1985), but with a small remnant population in Haiti (Turvey et al. 2008). Solenodons
70 inhabit the forests of Cuba and Hispaniola to elevations of 2000 m or more (Eisenberg
71 and Gozalez Gotera 1985; Ottewalder 1985) and shelter in caves, crevices, logs, and
72 extensive tunnel networks at a depth of >20 cm (Eisenberg and Gozalez Gotera 1985;
73 Ottewalder 1985). They subsist on small vertebrates and large invertebrates (Eisenberg
74 and Gozalez Gotera 1985; Ottewalder 1985). Solenodons are one of the few mammals
75 that are venomous (Turvey 2010; Folinsbee 2013), and some have argued that they may
76 use a form of echolocation (Eisenberg and Gould 1966).

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

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

94 Despite their unique status among mammals (Roca et al. 2004; Wible 2008), until
95 recently only a few studies had published on their DNA, these involving one or a few
96 specimens (Stanhope et al. 1998; Emerson et al. 1999; Allard et al. 2001; Waddell and
97 Shelley 2003; Roca et al. 2004). Likewise, solenodons are the only major mammalian
98 lineage for which the full length of the mitochondrial DNA genome has not been
99 sequenced. Both solenodon species are listed as endangered and declining in population
100 by the IUCN Red List of threatened species (IUCN 2013). Solenodons are currently
101 threatened with extinction due to deforestation, increasing human populations, predation
102 by introduced carnivores, and possibly by competition from introduced rodents
103 (Ottenwalder 1985; Ottenwalder 2001; Borroto-Paez 2009; IUCN 2013; Turvey et al.
104 2014). A better understanding of the genetics of solenodons would assist development of
105 conservation plans. In particular, regional differences in morphology have been reported

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

114

115 **Materials and Methods**

116

117 **Sample collection and DNA extraction**

118

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

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

141

142 **Mitogenome sequence generation, assembly and annotation**

143 For each of six individuals of *S. paradoxus*, a mitogenome was assembled
144 independently from Illumina WGS reads. From northern solenodon (*S. p. paradoxus*)
145 Spa-1, a total of 52,358,830 paired-end reads were generated, equating to approximately
146 13.09 Gb of sequence data. For five southern solenodons (*S. p. woodi*), an average of
147 151,783,327 paired-end reads were generated, equating to an average of 15.33 Gb of
148 sequence data for each individual. For each of the six individuals, a mitogenome was
149 assembled independently using two methods. First via a standard approach with MITO-
150 bim (Hahn et al. 2013) using the complete mitochondrial genome sequence of *Tapirus*
151 *indicus* (GenBank accession number NC_023838) as a reference seed. Separately, they

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

158 Complete mitogenome sequences were assembled separately for each individual.
159 Multiple sequence alignment with MAFFT (Katoh and Standley 2013) and a custom
160 variant calling script were employed to assess variation among mitogenomes. This
161 variation was confirmed by aligning reads from every individual to the assembly of every
162 other individual with Bowtie2 (Langmead and Salzberg 2012) and calling variants with
163 SAMtools and Bcftools (Li 2011). All solenodon mitogenomes reported in this study
164 were annotated and deposited in GenBank (accession numbers KU697358-KU697363).
165 Partial mitogenome sequences previously published for *S. cubanus* (Roca et al. 2004)
166 were likewise aligned to the consensus sequence with MAFFT (Katoh and Standley
167 2013), and variants were called.

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

175 for tRNA-Ser, expected at approximately the 11.5 kilobase mark, was not discovered
176 with this method. We used Exonerate (Slater and Birney 2005) with the ungapped model
177 to align tRNA coding genes of *Sorex araneus* (NCBI accession number NC_027963) to
178 the mitogenome of *Solenodon paradoxus woodi*. All of these aligned to the previously
179 discovered regions, with eight (for Ile, Gln, Ala, Asn, His, Ser, Leu, Pro) alignments
180 spanning their entire lengths, including the previously missed tRNA-Ser at bases 11632–
181 11690. The structure and polymorphisms of the mitogenomes and of mtDNA segments
182 were visualized using the software Circos (Krzywinski et al. 2009).

183 To assess amino acid substitutions, pairwise alignments of protein coding regions
184 of *S. paradoxus woodi* to *Talpa europaea* (NCBI accession number NC_002391), *Sorex*
185 *araneus* (NC_027963), *Felis catus* (NC_001700) and *Homo sapiens*
186 (GCA_000001405.20) were performed with the software Exonerate (Slater and Birney
187 2005) model coding2coding. Provean (Choi and Chan 2015) scores were calculated for
188 the substitutions. A stringent score cutoff of -4.1 was used to identify mutations most
189 likely to have effect on protein function; the complete list of these mutations, grouped by
190 genes, is presented in Supplementary Table S1.

191

192 **Phylogenetic analyses and molecular dating**

193 For phylogenetic analysis, nucleotide sequences were aligned using webPRANK
194 (Loytynoja and Goldman 2010) in EBI Web Services (McWilliam et al. 2013); alignment
195 output was visually inspected and edited using Mesquite v3.04
196 (<http://mesquiteproject.org>). Substitution models were determined using the Akaike
197 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

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

224 To place the solenodon into a phylogeny with other mammals, the solenodon
225 mitogenomes (without control region) were aligned to those of the mammals listed in
226 Supplementary Table S2. These included a shrew (Soricidae) and mole (Talpidae),
227 representing the families most closely related to the Solenodontidae, along with 52
228 additional taxa from Laurasiatheria. The mitogenome of the hedgehog, *Erinaceus*
229 *europaeus*, was excluded from the analysis due to previously reported (Mouchaty et al.
230 2000) significant deviation of the published mitogenome sequence from expected
231 nucleotide and amino acid compositions. For this analysis, a total of 15,624 aligned
232 nucleotide positions were analyzed after poorly aligned regions (mainly within 12S and
233 16S rRNA regions) were excluded. The software RAxML version 8.2.4 (Stamatakis
234 2014) was used to generate a maximum likelihood tree, using the GTR+I+G substitution
235 model and 1000 rapid bootstrap replicates.

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

244 randomly generated starting trees and birth-death process tree prior (Gernhard 2008). A
245 monophyly constraint was applied to non-eulipotyphlan taxa to minimize spurious dating
246 resulting from incorrect trees and to enforce the correct topology (Eulipotyphla sister to
247 all other laurasiatherians), which has been well established using nuclear data (Roca et al.
248 2004; Meredith et al. 2011). Posterior distributions were obtained by MCMC sampling
249 from at least 60,000,000 steps, with a discarded burn-in of at least 6,000,000; samples
250 were drawn every 1000 MCMC steps.

251

252 **Results**

253

254 The consensus mitogenome of the Hispaniolan solenodon was 16,457 bp, and
255 consists of the expected repertoire of 37 genes, including 13 that code for polypeptides,
256 22 for transfer RNA (tRNA) and the small and large subunits of ribosomal RNA (rRNA).
257 The length of the mitogenome was: 16,457 bp for Spa-K; 16,456 bp for Spa-L; 16,455 for
258 Spa-M, -N and -O; and 16,454 for Spa-1. Figure 2 represents gene and variant tracks
259 plotted with Circos (Krzywinski 2009). The five individuals of *S. paradoxus woodi*
260 exhibited little sequence variation (Figure 2). Spa-M, -N and -O were identical in
261 sequence. Between Spa-K and Spa-L there were 2 differences. And for Spa-M/N/O there
262 were 4 differences with Spa-K and 2 differences with Spa-L. By contrast, the
263 mitogenome of the individual of *S. paradoxus paradoxus* (Spa-1) exhibited 100
264 differences from the consensus of the 5 southern solenodons; and the mtDNA fragments
265 of *S. cubanus* showed an even greater density of differences (Figure 2). Coding genes
266 were discovered and annotated *de novo*. The relative positions and lengths of the genes in

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

275 The complete mitogenomes (including control regions) of Hispaniolan solenodons
276 were used to infer a phylogeny in BEAST. The 5 solenodons from southern Hispaniola
277 formed a clade with the northern Hispaniolan solenodon as an outgroup to the clade
278 (Figure 3A). Using two evolutionary rates (the fast rate of house mouse (Goios et al.
279 2007), 3.72×10^{-8} and the slow rate of humans (Soares et al. 2009), 1.665×10^{-8}
280 substitutions per site per year) as extremes, the point estimate for the divergence between
281 the two clades was estimated as 78,000 and 171,400 years, respectively. Using an
282 alignment of 12S and 16S rRNA regions for which *S. cubanus* sequence could be used as
283 an outgroup confirmed that the split in *S. paradoxus* was between the 5 southern and 1
284 northern Hispaniolan solendons (Figure 3B). In an attempt to increase the number of
285 northern individuals that could be analyzed, the control region sequences of the *S.*
286 *paradoxus* genomes were aligned with those previously published for three solenodons,
287 two of northern provenance (Loma de la Jagua; Figure 1) and one of unknown origin. In
288 a network analysis of the control region sequences, there was separation between the
289 northern (n=3) and southern (n=5) individuals (Figure 3C). For this dataset, F_{ST} between

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

293 Our solenodon mitogenomic sequences (excluding control region) were then
294 aligned with those of 54 other mammalian taxa (Supplementary Table S2). The
295 relationships inferred for the taxa are shown in Supplementary Figures S1 and S2, and
296 confirms that solenodons were basal to soricids and talpids. Relying on well-established
297 fossil calibration dates reported previously (Supplementary Table S3) (Meredith et al.
298 2011), we estimated that solenodons had diverged from other mammals ca. 78.2 Mya
299 (95% CI of 62.1-98.6 Mya). In this case, the estimated coalescence date for the *S.*
300 *paradoxus* sequences (north vs. south) was 0.60 (95% CI of 0.33-0.94) Mya.

301

302 Discussion

303

304 There may be a number of potential caveats when considering the relationships
305 among some taxa or the molecular dating: (1) there is a potential that some published
306 mitogenomic sequences may have unwittingly incorporated numts (Lopez et al. 1994);
307 (2) the analyses may be affected by different evolutionary rates across branches; (3) long
308 branch attraction may potentially affect relationships; (4) various factors may lead to
309 mito-nuclear incongruence, notably hybridization and sex-biased gene flow (Li et al.
310 2016). Given those potential caveats, it is noteworthy that using mammalian
311 mitogenomes, the solenodons are estimated to have diverged from other mammals ca.
312 78.2 Mya (95% CI of 62.1-98.6 Mya). This is remarkably close to the previous point

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

316 This time estimate would be consistent with a vicariant origin for solenodons,
317 since the proto-Antilles may have been connected to the North American continent some
318 70-80 Mya (Hedges 1996; Iturralde-Vinent and MacPhee 1999; Acton *et al.* 2000).
319 However, the end of the Cretaceous 65 Mya is marked by the impact of an asteroid at
320 nearby Chicxulub in the Yucatan (Iturralde-Vinent and MacPhee 1999). Small fossorial
321 mammals may be more likely to survive some catastrophic events (Andersen and
322 MacMahon 1985), and it is unlikely that islands of the proto-Antilles were completely
323 inundated by the tsunami caused by the shallow-sea impact (Bryant 2014). Nonetheless,
324 any life in the proto-Antilles would have faced catastrophic effects, especially since the
325 islands would have been even closer to the impact site than they are today (Hedges 1996).
326 Additionally, although some geologists have stated that the mountains of Hispaniola have
327 remained above sea level continuously since the Mesozoic (Donnelly 1992), others have
328 argued that every Caribbean island has been submerged at some point during the
329 Cenozoic (Iturralde-Vinent and MacPhee 1999), which if true would rule out a
330 completely vicariant origin for the solenodon lineage.

331 A Mesozoic divergence date versus other mammals does not necessarily imply a
332 vicariant origin, as the solenodon lineage may have diverged more recently from
333 mainland mammals that are now extinct. Some North American extinct taxa thought to be
334 related to *Solenodon*, such as *Centetodon* or *Apternodus* (Asher *et al.* 2002; Lopatin
335 2003), did not appear to be close to *Solenodon* when morphological datasets were

336 analyzed while constrained to a scaffold tree of known molecular relationships (Asher et
337 al. 2002; Roca et al. 2004). The recently extinct Nesophontidae (Whidden and Asher
338 2001; Rzebik-Kowalska and Woloszyn 2012; Orihuela 2014), a family of 11 species of
339 insectivores that also lived only in the islands of the West Indies but went extinct after
340 European settlement, also did not appear close to solenodons in that analysis. One issue
341 with a dispersal of solenodons into the Antilles is that they are present only in Cuba and
342 Hispaniola, with their origin in the two islands attributed to vicariance, since the
343 divergence time between the two species (Roca et al. 2004) occurred when the two
344 islands separated geologically (Iturralde-Vinent and MacPhee 1999). Since solenodons
345 are also not found in other Caribbean islands, they are unlikely to be natural dispersers.
346 *Nesophontes* appears to have more regularly dispersed across islands, as species were
347 present in Puerto Rico and the Cayman Islands for which dispersal is the only plausible
348 explanation (Whidden and Asher 2001). Recently, an analysis of endocranial morphology
349 has suggested similarities between *Solenodon* and *Nesophontes* (Orihuela 2014). If a
350 sister relationship is upheld between *Solenodon* and *Nesophontes*, one can hypothesize
351 that the ancestral form was more likely to resemble *Nesophontes*, for which there is
352 strong evidence for dispersal, and that the morphology of non-dispersing *Solenodon* may
353 be derived subsequent to arrival.

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

359 of the Cul de Sac Plain/Neiba Valley (see Figure 1) were placed into the subspecies *S. p.*
360 *paradoxus*. These differences were likely due to repeated separation of northern and
361 southern Hispaniola by a marine canal into north and south paleo-islands during the
362 Pliocene and Pleistocene (Ottewalder 2001). Our results are consistent with the
363 proposed subspecies division, since solenodons from the northern and southern
364 Dominican Republic (Figure 1) separate in a haplotype network (Figure 3C). The
365 southern Hispaniola solenodons appear to have much less genetic diversity than those in
366 the north (Figure 3C), which may be related to the sixfold larger size of the northern
367 paleo-island over the southern paleo-island (Figure 1) (Ottewalder 2001). The estimated
368 F_{ST} of 0.72 between northern and southern populations suggests a high degree of
369 differentiation between the two geographic groups. This division into two subspecies
370 would have needed confirmed by additional sampling and by the use of nuclear genetic
371 markers. As this manuscript was in the peer-review process, a broader survey of
372 solenodons across Hispaniola was reported (Turvey et al. 2016), based on 534 bp of the
373 control region and 411 bp of cytochrome b in 34 samples of *Solenodon paradoxus*. They
374 report the species to be genetically divided into northern, south-eastern and south-western
375 populations (the last being confined to the Massif de la Hotte in Haiti) (Turvey et al.
376 2016). They also report morphometric support for this subdivision using 110 specimens
377 (Turvey et al. 2016). Our results here thus confirm their finding that the northern and
378 south-eastern populations are distinctive, and that the south-eastern population has very
379 limited mitochondrial DNA diversity, and provide further support for the conservation
380 management of Hispaniolan solenodons as distinct regional subspecies or evolutionarily
381 significant units.

382

383 **Acknowledgements**

384

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

386

387 **Declaration of Interest Statement**

388

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

390

391 **Funding**

392

393 SJO, AK, and PD were supported by the Russian Ministry of Science Mega-grant

394 no.11.G34.31.0068.

395

396

397 **References**

398

399 Acton GD, Galbrun B, King JW. 2000. Paleolatitude of the Caribbean Plate since the

400 Late Cretaceous. In: Leckie RM, Sigurdsson H, Acton GD, Draper G, editors.

401 Proc. ODP, Sci. Results. College Station (TX): Ocean Drilling Program. p. 149-

402 173.

403 Allard MW, Baker SD, Emerson GL, Ottenwalder JA, Kilpatrick CW. 2001.

404 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 Zool 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, Chevreur 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.

- 470 Lopatin AV. 2003. A zalambdodont insectivore of the family Apternodontidae
471 (Insectivora, Mammalia) from the Middle Eocene of Mongolia. *Paleontol J.*
472 37:187-195.
- 473 Lopez JV, Yuhki N, Masuda R, Modi W, O'Brien SJ. 1994. Numt, a recent transfer and
474 tandem amplification of mitochondrial DNA to the nuclear genome of the
475 domestic cat. *J Mol Evol.* 39(2):174-90.
- 476 Loytynoja A, Goldman N. 2010. webPRANK: a phylogeny-aware multiple sequence
477 aligner with interactive alignment browser. *BMC Bioinformatics.* 11:579.
- 478 MacPhee RDE, Flemming C, Lunde DP. 1999. "Last occurrence" of the Antillean
479 insectivoran *Nesophontes*: new radiometric dates and their interpretation. *Amer*
480 *Mus Novit.* 3261:1-20.
- 481 MacPhee RDE, Novacek MJ. 1993. Definitions and relationships of Lipotyphla. In:
482 Szalay FS, Novacek MJ, McKenna MC, editors. *Mammal Phylogeny: Placentals.*
483 New York: Springer-Verlag. p. 13-31.
- 484 McDowell SB, Jr. 1958. The Greater Antillean insectivores. *Bull Amer Mus Nat Hist.*
485 115:113-214.
- 486 McKenna MC, Bell SK. 1997. *Classification of Mammals above the Species Level.* New
487 York: Columbia University Press.
- 488 McWilliam H, Li WZ, Uludag M, Squizzato S, Park YM, Buso N, Cowley AP, Lopez R.
489 2013. Analysis Tool Web Services from the EMBL-EBI. *Nucleic Acids Res.*
490 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.

- 537 Turvey ST, Meredith HMR, Scofield RP. 2008. Continued survival of Hispaniolan
538 solenodon *Solenodon paradoxus* in Haiti. *Oryx*. 42:611-614.
- 539 Turvey ST, Peters S, Brace S, Young RP, Crumpton N, Hansford J, Nunez-Mino JM,
540 King G, Tsalikidis K, Ottenwalder JA, et al. 2016. Independent evolutionary
541 histories in allopatric populations of a threatened Caribbean land mammal.
542 Diversity Distrib. DOI: <http://dx.doi.org/10.1111/ddi.12420>
- 543 Waddell PJ, Shelley S. 2003. Evaluating placental inter-ordinal phylogenies with novel
544 sequences including RAG1, gamma-fibrinogen, ND6, and mt-tRNA, plus
545 MCMC-driven nucleotide, amino acid, and codon models. *Mol Phylogenet Evol.*
546 28:197-224.
- 547 Wheeler DL, Church DM, Federhen S, Lash AE, Madden TL, Pontius JU, Schuler GD,
548 Schriml LM, Sequeira E, Tatusova TA et al. 2003. Database resources of the
549 National Center for Biotechnology. *Nucleic Acids Res.* 31:28-33.
- 550 Whidden HP, Asher RJ. 2001. The origin of the Greater Antillean insectivorans. In:
551 Woods CA, Sergile FE, editors. *Biogeography of the West Indies: Patterns and*
552 *Perspectives*. 2nd ed. Boca Raton (FL): CRC Press. p. 237-252.
- 553 Wible JR. 2008. On the cranial osteology of the Hispaniolan Solenodon, *Solenodon*
554 *paradoxus* Brandt, 1833 (Mammalia, Lipotyphla, Solenodontidae). *Ann Carnegie*
555 *Mus.* 77:321-385.
- 556 Woods CA, Sergile FE. 2001. *Biogeography of the West Indies: Patterns and*
557 *Perspectives*. Boca Raton (FL): CRC Press.
- 558
- 559

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 (Ottewald 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

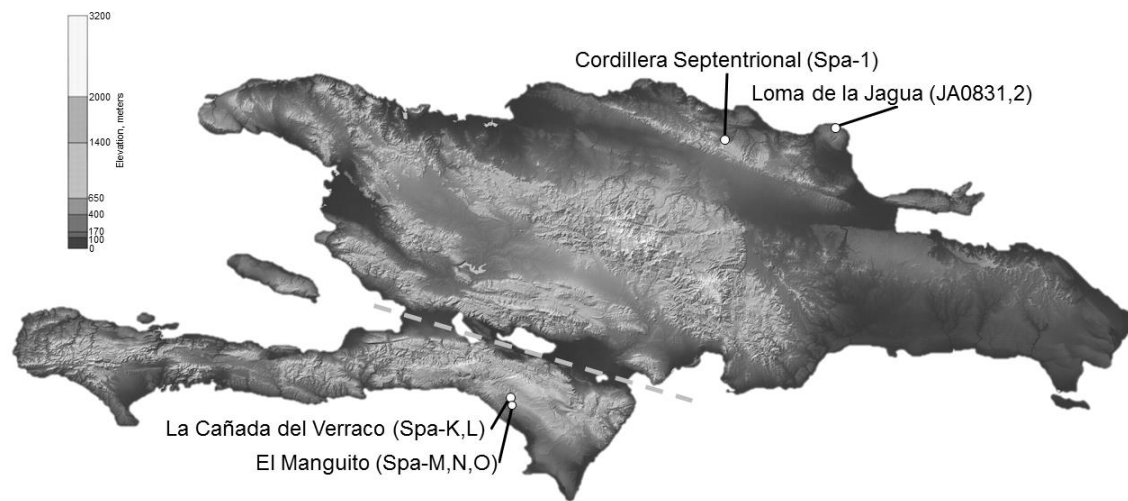
606 sequences were combined in a 389 bp alignment with previously generated control region
607 sequences from two specimens from the northern Dominican Republic, and one of
608 unknown provenance. The network shows each difference as a hatch mark (all were
609 substitutions, there were no indels). The 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

616

[[Figure 1]]

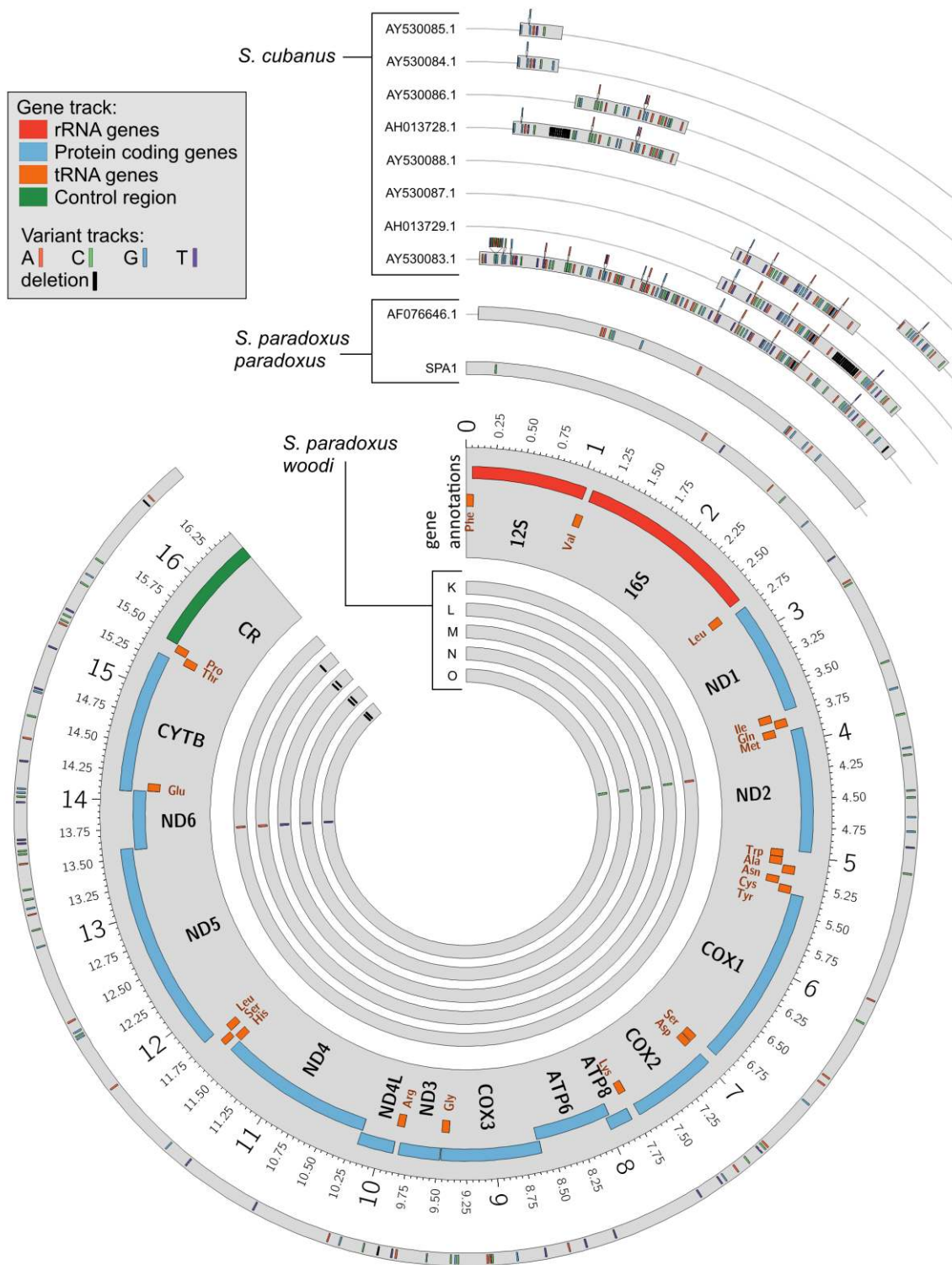
617

618



619

[[Figure 2]]



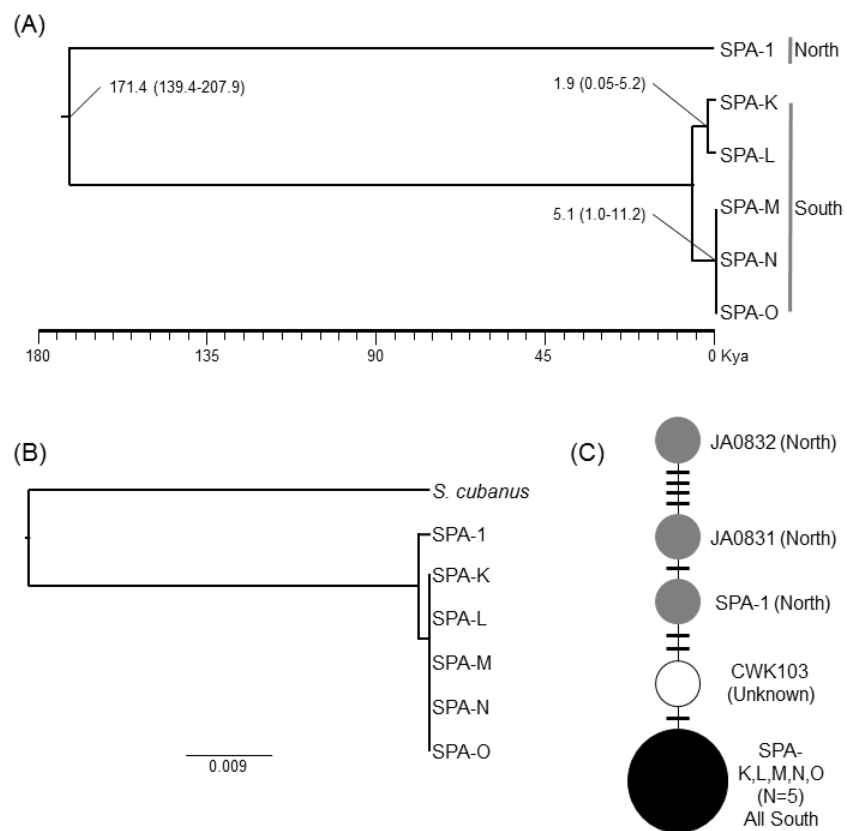
620

621

[[Figure 3]]

622

623



624 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 18° 09' 9.64" W 71° 43' 12.0"	This study	
SPA-L	South	Pedernales	La Cañada del Verraco	M	Wild	1020	N 18° 09' 9.64" W 71° 43' 12.0"	This study	
SPA-M	South	Pedernales	El Manguito - 1	M	Wild	1270	N 18° 06' 36.6" W 71° 43' 3.58"	This study	
SPA-N	South	Pedernales	El Manguito - 1	F	Wild	1420	N 18° 06' 36.6" W 71° 43' 3.58"	This study	
SPA-O	South	Pedernales	El Manguito - 2	F	Wild	1120	N 18° 07' 6.5" W 71° 43' 14.7"	This study	
SPA-1	North	Espailat	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

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

626

627

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

Supplementary Material

Authors: Adam L. Brandt^{1,2+}, Kirill Grigorev⁴⁺, Yashira M. Afanador-Hernández⁴, Liz A. Paulino⁵, William J. Murphy⁶, Adrell Núñez⁷, Aleksey Komissarov⁸, Jessica R. Brandt¹, Pavel Dobrynin⁸, J. David Hernández-Martich⁹, Roberto María⁷, Stephen J. O'Brien^{8,10}, Luis E. Rodríguez⁵, Juan C. Martínez-Cruzado⁴, Taras K. Oleksyk^{4*} and Alfred L. Roca^{1,2,3*}

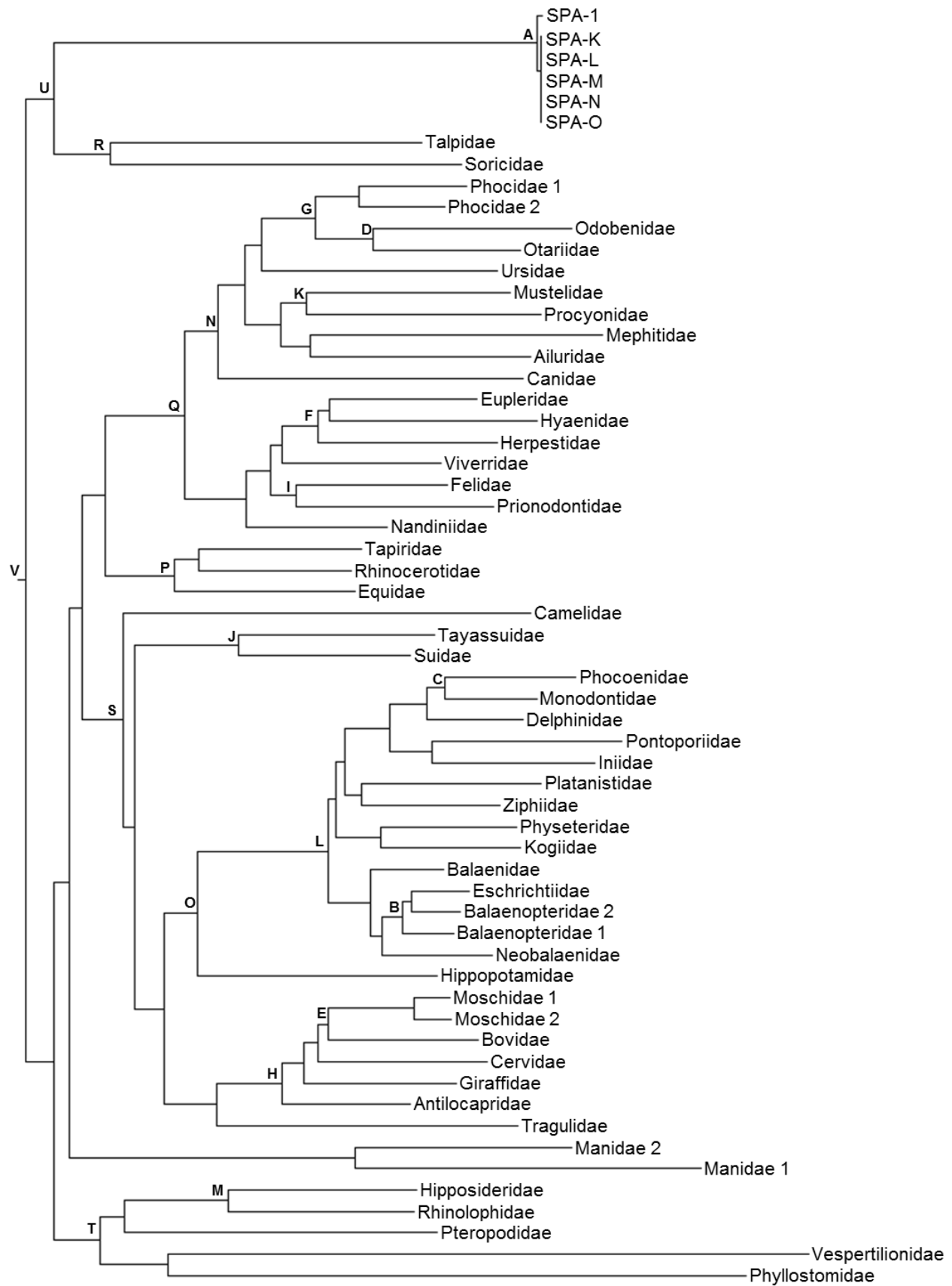
⁺Equal contributors

^{*}Corresponding authors: roca@illinois.edu; taras.oleksyk@upr.edu

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



0.08

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).

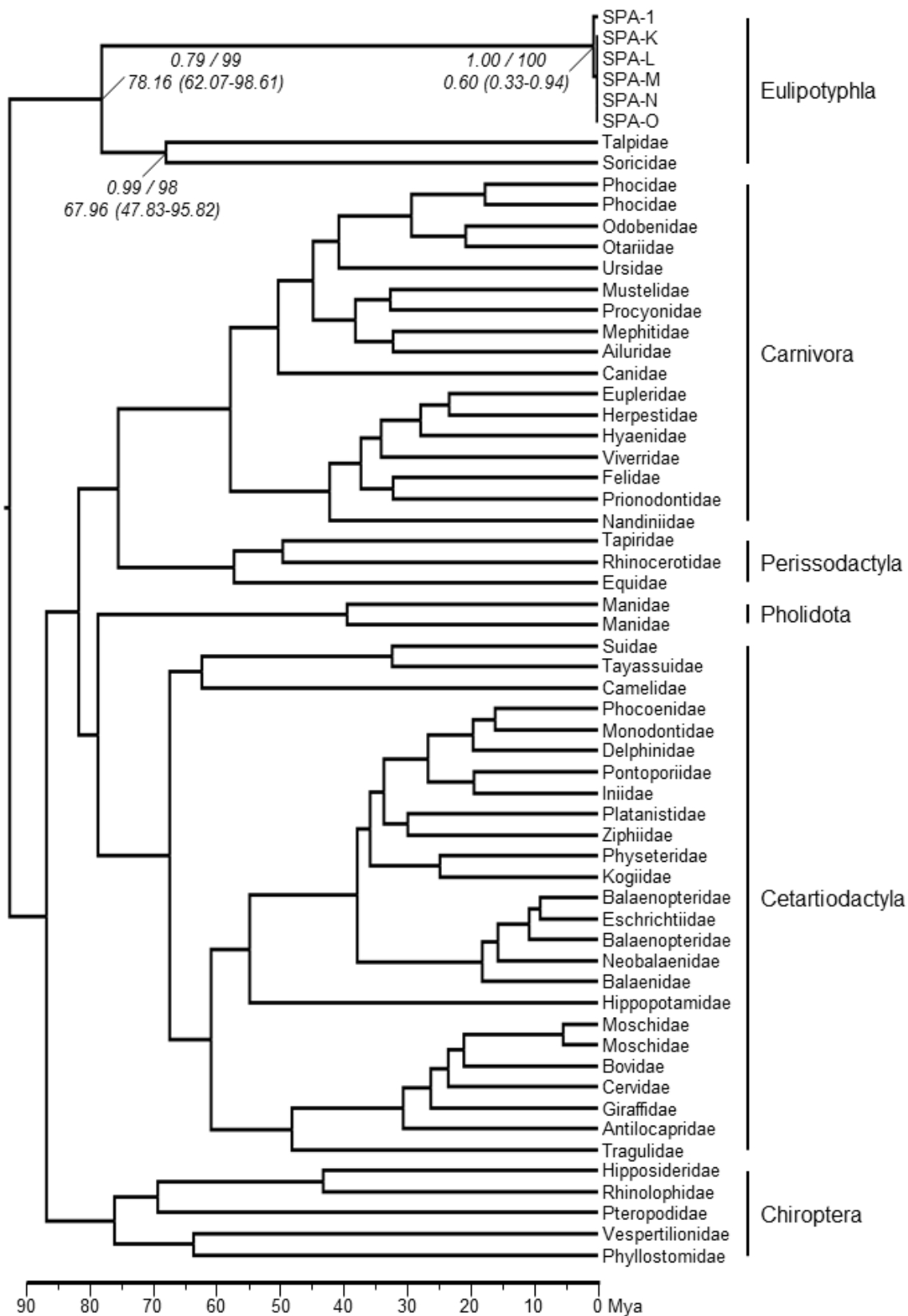


Table S1. Radical mutations in mitochondrial protein coding genes of *S. paradoxus* in comparison to those of reference species. Mutations are shown in this format: position, amino acid in the reference mitogenome, amino acid in *S. paradoxus*. Scores were calculated with PROVEAN using a stringent threshold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

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

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

Group	Node	Fossil*	Estimated (95% CI)	Posterior probability	Bootstrap node support (%)
Base of <i>S. paradoxus</i>	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.