

Molecular phylogenetic analysis of nuclear genes suggests a Cenozoic over-water dispersal origin for the Cuban solenodon

著者別名	北 将樹
journal or publication title	Scientific reports
volume	6
page range	31173
year	2016-08
権利	(C) The Author(s) 2016 This work is licensed under a Creative Commons Attribution 4.0 International License. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit http://creativecommons.org/licenses/by/4.0/
URL	http://hdl.handle.net/2241/00143961

doi: 10.1038/srep31173

SCIENTIFIC REPORTS

OPEN

Molecular phylogenetic analysis of nuclear genes suggests a Cenozoic over-water dispersal origin for the Cuban solenodon

Jun J. Sato¹, Satoshi D. Ohdachi², Lazaro M. Echenique-Diaz³, Rafael Borroto-Páez⁴, Gerardo Begué-Quiala⁵, Jorge L. Delgado-Labañino⁶, Jorgelino Gámez-Díez⁶, José Alvarez-Lemus⁷, Son Truong Nguyen⁸, Nobuyuki Yamaguchi⁹ & Masaki Kita¹⁰

Received: 01 April 2016
Accepted: 13 July 2016
Published: 08 August 2016

The Cuban solenodon (*Solenodon cubanus*) is one of the most enigmatic mammals and is an extremely rare species with a distribution limited to a small part of the island of Cuba. Despite its rarity, in 2012 seven individuals of *S. cubanus* were captured and sampled successfully for DNA analysis, providing new insights into the evolutionary origin of this species and into the origins of the Caribbean fauna, which remain controversial. We conducted molecular phylogenetic analyses of five nuclear genes (*Apob*, *Atp7a*, *Bdnf*, *Brca1* and *Rag1*; total, 4,602 bp) from 35 species of the mammalian order Eulipotyphla. Based on Bayesian relaxed molecular clock analyses, the family Solenodontidae diverged from other eulipotyphlan in the Paleocene, after the bolide impact on the Yucatan Peninsula, and *S. cubanus* diverged from the Hispaniolan solenodon (*S. paradoxus*) in the Early Pliocene. The strikingly recent divergence time estimates suggest that *S. cubanus* and its ancestral lineage originated via over-water dispersal rather than vicariance events, as had previously been hypothesised.

As a natural laboratory, the Caribbean islands have garnered considerable attention worldwide because of their unique biogeographic features that have contributed to our understanding of organismal evolution^{1,2}. In particular, the species assembly rule of Caribbean biota has been a subject of intensive research and discussion. Since the early 20th century, there have been long-standing debates between vicariance and dispersal mechanisms of origin for the Caribbean species assembly^{2–10}. The most heated argument has been whether land bridges between the Caribbean islands and nearby continents in the Cenozoic Era contributed to the colonisation of organisms onto the islands. Previous geological studies proposed the GAARlandia hypothesis, in which a land bridge was formed between the Greater Antilles (Cuba, Hispaniola, Jamaica and Puerto Rico) and South America via the Aves Ridge (AR) around 33–35 Ma (mega annum)^{7,11}. Supporting evidence for GAARlandia has been obtained for various species^{9,12–16}; however, these studies did not provide exclusive support for this hypothesis, either because dispersal can also explain the results^{9,12,13,15}, or the data do not precisely coincide with the time scale of GAARlandia (33–35 Ma)^{14,16}. However, most empirical studies on vertebrate species have supported the over-water dispersal hypothesis for the origin of West Indian biota^{4,6,10,17,18}, rather than overland dispersal and subsequent vicariance

¹Laboratory of Animal Cell Technology, Faculty of Life Science and Technology, Fukuyama University, Higashimura-cho, Aza, Sanzo, 985, Fukuyama 729-0292, Japan. ²Institute of Low Temperature Science, Hokkaido University, Kita-19 Nishi-8, Kita-ku, Sapporo 060-0819, Japan. ³Environmental Education Center, Miyagi University of Education, Aramaki Aza-Aoba, Aoba-ku, Sendai 980-0845, Japan. ⁴Sociedad Cubana de Zoología, CP 11900, Boyeros, La Habana, Cuba. ⁵Unidad Presupuestada Parque Nacional Alejandro de Humboldt (CITMA), Calle Abogado 14 e/ 12 y 13 Norte, Guantánamo 95200, Cuba. ⁶Estación Ecológica La Melba, Unidad Presupuestada Parque Nacional Alejandro de Humboldt, CITMA-Guantánamo, Cuba. ⁷Centro de Inspección y Control Ambiental (CICA), Ministerio de Ciencia, Tecnología y Medio Ambiente (CITMA), Cuba. ⁸Institute of Ecology and Biological Resources, Vietnam Academy of Science and Technology, 18 Hoang Quoc Viet Street, Cau Giay, Hanoi, Vietnam. ⁹Department of Biological and Environmental Sciences, College of Arts and Sciences, Qatar University, PO Box 2713 Doha, Qatar. ¹⁰Graduate School of Pure and Applied Sciences and Tsukuba Research Center for Interdisciplinary Materials Science (TIMS), University of Tsukuba 1-1-1 Tennodai, Tsukuba 305-8571, Japan. Correspondence and requests for materials should be addressed to J.J.S. (email: jsato@bt.fubt.fukuyama-u.ac.jp)

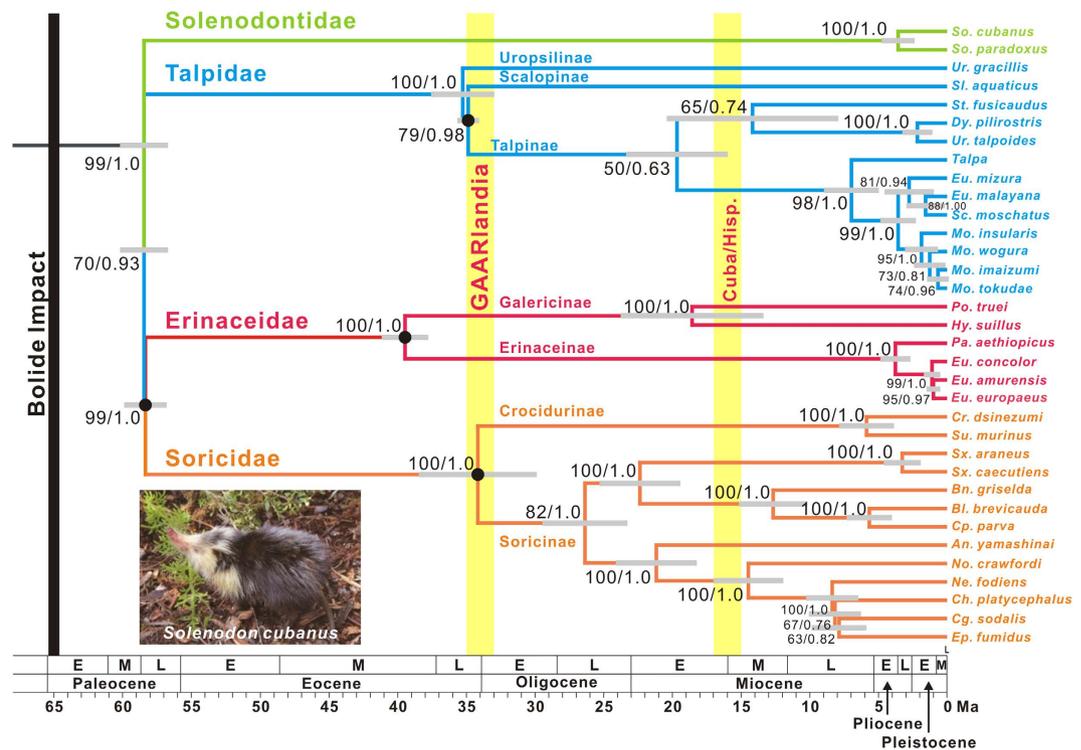


Figure 1. Bayesian chronogram of the order Eulipotyphla. The divergence times were estimated from the BEAST analyses of five nuclear genes (*Apob*, *Atp7a*, *Bdnf*, *Brca1* and *Rag1*; total, 4,602 bp). The maximum likelihood (ML) and Bayesian inference (BI) analyses both supported the same topology, as shown. The colours of each branch show four families in this order (green, Solenodontidae; blue, Talpidae; red, Erinaceidae; orange, Soricidae). Numbers near each node show bootstrap supports (left) and posterior probabilities (right) from the ML and BI analyses, respectively. The geological time scale for the Cenozoic era (65 Ma–present) is shown at the bottom and several geological events are marked on the chronogram by black or yellow thick vertical lines (Bolide Impact at 65 Ma, GAARlandia at 33–35 Ma, and geological separation between Cuba and Hispaniola [Cuba vs Hisp.] at ~16 Ma). The letters E, M and L in the time scale indicate Early, Middle and Late, respectively. The 95% credibility intervals for each divergence time estimate are represented by the grey horizontal bars. The nodes constrained for dating are represented by the black circles on each node. For details of the calibration settings, see the Methods section. The image of *Solenodon cubanus* was photographed by one of the authors (L.M.E).

in GAARlandia. In addition, a vicariant event unrelated to the GAARlandia hypothesis has been suggested for freshwater cichlid fish species between the Greater Antilles and the Yucatan Peninsula in Central America¹⁹. Therefore, the colonisation mechanism of the Caribbean biota remains elusive, and a generalisation of the species establishment process is unresolved.

The Cuban solenodon or almiquí (*Solenodon cubanus*) is a large insectivore (Fig. 1) and an endangered species isolated on the island of Cuba²⁰. It is listed as Endangered B1ab(iii,v) on the IUCN Red List, indicating that its distribution is severely fragmented in several limited locations and the population continues to decline^{21,22}. Moreover, the capture record of live solenodons is extremely poor, making this species one of the most enigmatic mammals worldwide. Therefore, basic biological data are required for their conservation. Together with the Hispaniolan solenodon (*S. paradoxus*), these two extant species constitute the family Solenodontidae in the order Eulipotyphla²³. Because the fossil record of Solenodontidae, especially the genus *Solenodon*²⁴, is limited, the evolution and systematic relationships within Eulipotyphla cannot be accurately investigated without access to living solenodon individuals. Clarifying the establishment process of this species in Cuba could contribute to resolving the Caribbean faunal assembly rule.

There have been only a few investigations into the origin of the Cuban solenodon. The divergence time between the Cuban and Hispaniolan solenodons was estimated to be 25 Ma (95% credibility interval [CI] = 16–38 Ma) based on a molecular chronological analysis of mitochondrial DNA (mtDNA) using old museum specimens²⁵. Roca *et al.*²⁵ indicated that there is consistency between the molecular estimate and geological evidence that the separation between Cuba and Hispaniola occurred at 25–27 Ma⁷ and suggested that the two solenodon species could have originated via a vicariance event. However, there are two problems with their conclusion. First, the geological evidence has been updated and it was inferred that the separation between these two islands occurred ~16 Ma¹¹, which does not support their molecular estimate of 25 Ma, although the 95% CI marginally covers 16 Ma²⁵. Second, only mtDNA was examined in their study on the Cuban solenodon, despite using multiple nuclear DNA (nucDNA) genes for other eulipotyphlan species. Dating old divergences with mtDNA is known

to be problematic because of the rapid evolutionary rate of mtDNA²⁶. It is well known that genetic analyses of old museum samples are difficult due to the degraded state of the DNA samples, especially for nucDNA genes²⁷. Under such conditions, only mtDNA with large copy numbers have been examined. The more slowly evolving nucDNA would provide better insight into the time scale of solenodon evolution; however, to analyse nucDNA, fresh tissue samples are necessary. Recently, we successfully obtained fresh tissue samples of the rare Cuban solenodon (see Methods), which allowed for nucDNA sequence analyses.

For the phylogenetic and chronological analyses in this study, we used five exon regions of slowly evolving nucDNA genes (*Apob*, apolipoprotein B; *Atp7a*, ATPase, Cu⁺⁺ transporting, alpha polypeptide; *Bdnf*, brain-derived neurotrophic factor; *Brcal*, breast cancer 1, early onset; *Rag1*, recombination activating gene 1), all of which have been used in previous molecular systematic studies of mammals^{25,28–31}. Our analysis of the phylogenetic relationships and divergence times using a molecular supermatrix composed of the five nuclear gene exon sequences provided evidence supporting over-water dispersal, rather than vicariance, for the origin of the family Solenodontidae in the Paleocene and the Cuban solenodon in the Early Pliocene.

Results and Discussion

Origin of Solenodontidae. The supermatrix phylogenetic analyses of five exon regions of nucDNA genes (total, 4,602 bp) demonstrated that Solenodontidae diverged first among the four families of Eulipotyphla (Solenodontidae, Erinaceidae, Soricidae and Talpidae²³; Fig. 1 and Supplementary Fig. S1; bootstrap proportion = 70, posterior probability = 0.93), followed by Talpidae, with Erinaceidae and Soricidae forming a clade. This is congruent with one previous study²⁵, but contradicts another supporting a phylogenetic affinity between Soricidae and Talpidae using 26 nuclear genes (total, 35,603 bp)³¹. Despite the difference in some of the inter-familial relationships, the basal position of Solenodontidae was robust for the examined gene sequences among these studies, including ours. Conversely, previous studies might have encountered the long-branch attraction (LBA)³² problem because only one or two representative species of each family separated by long branches were examined. For the phylogenetic and chronological analyses of this study, we sampled all four families in Eulipotyphla and all of the subfamilies in each family, where available (Erinaceinae and Galericinae in Erinaceidae, Crocidurinae and Soricinae in Soricidae, and Scalopinae, Talpinae, and Uropsilinae in Talpidae; Supplementary Table S1). Using this sampling procedure, we conducted analyses using all of the major lineages in Eulipotyphla at the subfamilial level, making the phylogenetic analyses less vulnerable to the LBA problem for inter-familial relationships. Therefore, our analyses using all subfamilial lineages largely strengthened the hypothesis of the basal branching of Solenodontidae and the close affinity of Erinaceidae and Soricidae. The phylogenetic inconsistency between this study and that by Meredith *et al.*³¹ and the very short branch lengths among the four major families in the chronogram (Fig. 1) suggest that rapid cladogenesis within a short evolutionary period was involved in these divergences; thus, further data are required for complete resolution of the inter-familial phylogenetic relationships.

The vicariance and dispersal hypotheses have been intensively debated as the origin of Solenodontidae. In recent molecular phylogenetic studies, the emergence of the ancestral lineage of Solenodontidae was inferred to have occurred in the late Cretaceous period (76 Ma [95% CI = 72–81 Ma] in Roca *et al.*²⁵; 77.3 Ma [95% CI = 70.7–85.8 Ma] in Meredith *et al.*³¹), which is chronologically consistent with the previous supposition that vicariance and subsequent restriction to the proto-Antilles, located between North and South America (near current Central America), in the Late Cretaceous period^{6,33} generated the ancestral lineage of Solenodontidae¹⁰. This is not a new hypothesis, since Rosen³³ depicted the hypothetical relative movements of the proto-Antilles northeastward from the region of the present Central America. However, palaeogeographical evidence is insufficient to support the proto-Antillean vicariance hypothesis⁶. Moreover, it seems implausible that ancestral solenodon populations could have survived the devastating impacts of the gigantic bolide on the Yucatan Peninsula (Chicxulub) at 65 Ma^{3,4,6,11,17}, which was also a concern of Roca *et al.*²⁵ in their discussion of the late Cretaceous divergence of Solenodontidae.

In this study, we tested the hypothesis of a Late Cretaceous origin of Solenodontidae with multiple fossil calibrations within Eulipotyphla³⁴, which was accomplished using our extensive sampling strategy. Using only eulipotyphlan fossil calibrations, we obtained an estimate of 58.6 Ma (95% CI = 57.3–60.8 Ma) for the age of the most recent common ancestor (TMRCA) of Eulipotyphla (Fig. 1 and Supplementary Table S2; strategy [1]), which is later than the age of the bolide impact (65 Ma; Fig. 1). This estimate is also incongruent with the GAARlandia land bridge in the Eocene-Oligocene boundary period (33–35 Ma; Fig. 1). Therefore, it supports the possibility that early Cenozoic dispersal after the bolide impact, not Late Cretaceous proto-Antillean vicariance, produced the ancestral lineage of Solenodontidae as previously suggested⁴, although incomplete fossil records could promote underestimation of the divergence times. Supporting this, the North American origin of the over-water dispersal was suggested based on fossil evidence³⁵. Such an over-water dispersal origin of extinct North American soricomorph relatives has been considered a possible explanation for the current relictual distribution of Solenodontidae⁷.

Furthermore, Cenozoic intra-ordinal divergence after the Cretaceous-Cenozoic boundary (65 Ma) is a common trend in mammalian diversification³¹. Interestingly, Phillips³⁶ re-analysed the dataset of Meredith *et al.*³¹, taking into consideration the calibration and evolutionary rate errors among the major mammalian lineages, and produced a divergence time estimate for the TMRCA of Eulipotyphla, which is consistent with this study (59.1 Ma). Geologists have stated that the Greater Antilles was submerged until the Middle Eocene (ca 45 Ma)^{7,37}, and early mammalian fossils in the region date back to the Oligocene²², which is incongruent with the hypothesis of the proto-Antillean vicariance in the late Cretaceous period. However, it is also inconsistent with our estimate for the origin of ancestral Solenodontidae (58.6 Ma [95% CI = 57.3–60.8 Ma]). Currently, the idea that the ancestor of Solenodontidae originated from another location by dispersal no earlier than 58.6 Ma (after the emergence of the Greater Antilles) and “relictualised” might be the best hypothesis explaining the absence of fossil traces of the major continental groups in these islands that would otherwise have dispersed through the land bridge⁷.

Nevertheless, further independent molecular and palaeontological evidence or geological updates are necessary for more precise dating of the origin of Solenodontidae. Old divergences estimated in two previous studies may have been due to the use of completely or mostly non-eulipotyphlan fossils as calibration points in the estimation of divergence times among eulipotyphlan lineages^{25,31}, as discussed in Phillips³⁶. A recently published study also indicated a Late Cretaceous origin for Solenodontidae (78.7 Ma [95% CI = 62.1–98.6 Ma])³⁸. However, this study used the same fossil constraints as those of Meredith *et al.*³¹. Therefore, the time estimate of the Late Cretaceous age might be another example of overestimation.

Origin of the Cuban solenodon. The Cuban solenodon formed a clade with the Hispaniolan solenodon (Fig. 1). This relationship was highly supported by the bootstrap value (100%) and posterior probability (1.0) in the maximum-likelihood (ML) and Bayesian-inference (BI) analyses, respectively (Fig. 1 and Supplementary Fig. S1). Although we used various calibration standards (see Methods), the divergence times between the Cuban and Hispaniolan solenodons estimated by the Bayesian relaxed molecular clock analysis were robust to the difference in calibrations, demonstrating that the two solenodons diverged in the Early Pliocene (3.7–4.8 Ma [95% CI = 2.6–6.4 Ma]; Supplementary Table S2). This estimate is incongruent with the previous estimate (25 Ma [95% CI = 16–38 Ma])²⁵, which was used to suggest that the geological separation between Cuba and Hispaniola (25–27 Ma)⁷ would have led to the vicariance event between these species. Despite using 76 Ma for the basal divergence of the eulipotyphlan clade to compare the time scales between previous studies and our results (strategies [2] and [3]; see Methods), the basal divergence within Solenodontidae (*S. cubanus* versus *S. paradoxus*) was still estimated to have occurred in the Early Pliocene (4.6–4.8 Ma [95% CI = 3.1–6.4 Ma]; Supplementary Table S2), strikingly different from the previous estimate (25 Ma [95% CI = 16–38 Ma])²⁵.

It could be argued that the use of mtDNA caused overestimations in the divergence times of Roca *et al.*²⁵. Since mtDNA evolves much more rapidly than nucDNA, the saturation of substitutions becomes problematic when using mtDNA to estimate divergence times among distantly related species²⁶. Saturation leads to underestimation of substitutions, resulting in overestimation of divergence times among younger lineages than the calibration points³⁹. This could be the case for the previous mtDNA dating analysis²⁵, in which the use of mtDNA was unavoidable because of the old museum samples of *S. cubanus* used for the genetic analyses. Therefore, the younger estimate obtained using the more slowly evolving nucDNA in this study is considered to be a reasonable time estimate and we conclude that the Cuban solenodon originated in the Early Pliocene. It should be noted that although there is a two-fold difference in the branch lengths between the Cuban and Hispaniolan solenodon lineages in the ML tree topology (Supplementary Fig. S1), it does not pose a major problem because of the robustness of the time estimate obtained using our drastically different calibration strategy. The difference between the Cuban and Hispaniolan solenodons is only 1.2% (56 of 4602 bp). This small difference might have stochastically affected the branch length variation.

The Early Pliocene origin of the Cuban solenodon supports over-water dispersal from Hispaniola as the mechanism for its origin on the island of Cuba for several reasons. First, there have been no geological reports that imply land connections between Cuba and Hispaniola later than the Early to Middle Miocene (16 Ma)¹¹. Second, since the unidirectional ocean current around the Caribbean Sea flows northwestward from South America⁴, it is reasonable to assume that the Cuban lineage dispersed from Hispaniola, not the opposite. However, from a geological perspective, the present ocean surface current patterns have been suggested to be a feature of the last 4 million years⁷. Based on this argument, the over-water dispersal hypothesis has been criticised because the divergence was mostly dated to be >4 Ma^{4,7}. Conversely, the estimate obtained in this study (3.7–4.8 Ma) is not contradicted by the geological evidence that the current ocean circulation was initiated in the Pliocene period⁷. Third, a number of studies have supported the over-water dispersal hypothesis as the origin of the West Indian biota, especially non-volant mammals, reptiles and amphibians^{4,6,17,18}. Traditional examples for the over-water dispersal hypothesis have been provided by immunological-distance analysis^{5,6,17}, whereby most of the origins of West Indian vertebrates are not concentrated in a particular period, implying little or no relation to specific geological land bridges. A recent empirical study using a museumomics approach with next-generation sequencing technology supported the over-water dispersal of hutias (capromyine rodents with body sizes similar to solenodons) from South America⁴⁰. In that study, Hispaniola was suggested as a source for the later diversification of hutias in the Greater Antilles and Bahamas, which is in agreement with the dispersal direction inferred for solenodons in this study. Additional supportive evidence for the over-water dispersal origin from other islands, continents, and even from Caribbean islands has accumulated recently for a variety of taxonomic groups in the West Indies, including *Poitea* and *Erithalia* plants⁴¹, pupfishes⁴², eleutherodactyline frogs⁴³, *Anolis* lizards⁴⁴, alsophiine racer snakes⁴⁵, extinct oryzomyine rodents⁴⁶, short-faced bats⁴⁷ and platyrrhine primates⁴⁸. Therefore, our case study of *Solenodon* provides additional support for the dispersal origin of Caribbean vertebrates. All available evidence in recent molecular phylogenetic studies has shown that smaller invertebrates and a few vertebrates support the hypotheses of GAARlandia (*Selenops* spiders¹², butterflies¹⁵, orb-weaving spider¹⁶, cichlid fishes¹⁴, toads¹³) and other geological vicariance events (crickets⁴⁹), while most vertebrates with relatively larger body sizes support the over-water dispersal as a major role in their origins, suggesting that the biological properties of species could have affected the over-water dispersal processes of West Indian biota. This is consistent with a previous study that found that more distant island voles are larger in size than mainland or less distant island voles⁵⁰. This does not mean that smaller species were mostly affected by geological vicariance events, but rather that there may have been few survivors among smaller species in the high-risk passage of over-water dispersal.

Methods

Sample collection. The field survey was conducted in Alejandro de Humboldt National Park in the Republic of Cuba from March to April 2012, resulting in the capture of seven live Cuban solenodon individuals (Permit No. 03/2012 from the Cuban Environment Agency to collect DNA samples and export them to Japan, document

available upon request). After collecting the data, all individuals were carefully released into their habitat. All methods followed the approved guidelines provided by the Mammal Society of Japan. Including these newly captured seven Cuban solenodons, the nucleotide sequences of five exon regions of nucDNA genes, *Apob*, *Atp7a*, *Bdnf*, *Brca1* and *Rag1*, were examined for 35 species in the order Eulipotyphla (Supplementary Table S1). For the phylogenetic and chronological analyses, we obtained novel sequences from our experiments or downloaded previously determined sequences from the DDBJ/EMBL/GenBank international DNA databases (Supplementary Table S1). Three species in the order Carnivora (the Japanese weasel, *Mustela itatsi*, the ermine or stoat, *Mustela erminea*, and the small Asian mongoose, *Herpestes auro-punctatus*), one species in Scandentia (the northern tree shrew, *Tupaia belangeri*) and one species in Rodentia (the house mouse, *Mus musculus*) were used as outgroups in the phylogenetic analyses. Eulipotyphla and Carnivora were classified into the superorder Laurasiatheria, while the other Scandentia and Rodentia were classified into Euarchontoglires, the sister clade of Laurasiatheria³¹. These outgroups span taxonomically diverse groups and are expected to work well to fix the eulipotyphlan root.

Experimental techniques. Total genomic DNA was extracted from a piece of ear tissue using a standard phenol–chloroform method. All PCR reactions were performed with the KOD-Plus-NEO DNA Polymerase Kit (Toyobo, Osaka, Japan) in an automated thermal cycler (Life Touch thermal cycler, Bioer Technology, Hangzhou, China) using the following conditions: a 3-min denaturation period at 94 °C, 30 cycles of denaturation at 94 °C for 30 s, annealing at 50 °C for 30 s, extension at 68 °C for 90 s and a 10-min extension period at 68 °C. The PCR reaction mixture consisted of KOD-Plus-NEO buffer, 1.5 mM MgSO₄, 0.2 mM of each dNTP, 0.3 μM of forward and reverse primers for each gene (Supplementary Table S3), 1.0 U of KOD-Plus-Neo DNA polymerase and 0.1–0.2 μg of template total genomic DNA to a total volume of 50 μL. Nested PCR reactions were performed for *Apob*, *Brca1* and *Rag1* using each first and second PCR primer pair in two PCR steps, while non-nested reactions for *Atp7a* and *Bdnf* used only one pair of primers in one PCR step (Supplementary Table S3). In the second PCR, a 1-μL aliquot of each first PCR product was used as a template. The PCR conditions were the same as the first PCR. We sequenced the ca. 500–700 bp PCR products using the Big Dye Terminator Cycle Sequencing Kit ver. 3.1 (Thermo Fisher Scientific, Tokyo, Japan), followed by the ABI3130 automated sequencer (Applied Biosystems, Tokyo, Japan). All experimental protocols were approved by the Animal Care and Use Committee of Fukuyama University.

Construction of the molecular supermatrix. Multiple alignments of the determined sequences were conducted for each locus through the program MUSCLE implemented in MEGA ver. 6⁵¹. Then, we combined alignments of each locus into a single data matrix (supermatrix) containing 40 species and five nuclear gene sequences (*Apob*, 1,181 bp; *Atp7a*, 673 bp; *Bdnf*, 573 bp; *Brca1*, 1,071 bp; and *Rag1*, 1,104 bp; total, 4,602 bp). The supermatrix was composed of 230,100 taxon-by-character cells, where the number of missing cells was 73,351 (31.9%). Missing data were not relevant to the phylogenetic resolution, as the most important point in phylogenetic design is the informativeness of the collected data^{52,53}. Even data matrices with >60% missing cells can provide a well-resolved phylogeny if there are informative sequences in the data matrix⁵².

Phylogenetic analyses. We adopted two probabilistic phylogenetic inference approaches, ML and BI, using the constructed molecular supermatrix. Before conducting these phylogenetic analyses, we inferred the best-fit partitioning scheme and substitution models using PartitionFinder ver. 1.1.1⁵⁴. Partitions were pre-defined by genes and codons (15 partitions; five genes and three codon positions), and the best partitioning scheme was explored using the greedy search and determined by the Bayesian information criterion. The best partitioning scheme supported two subsets of partitions, where SYM+G was selected for the first and second codon positions of the *Bdnf* and *Rag1* genes, while GTR+I+Γ was used for the other 11 partitions. The ML phylogenetic analysis was conducted in GARLI ver. 0.96⁵⁵ using these combined substitution models. The ML tree was inferred through heuristic searches consisting of five independent runs of the genetic algorithm, each performing 2 × 10⁴ generations of mutation–selection–reproduction cycles with the starting tree generated from the ML stepwise–addition–sequence option. We followed the default option for the other parameters. Nodal support was evaluated using nonparametric bootstrap analyses (100 replicates).

The BI phylogenetic analysis was conducted in MrBayes ver. 3.1.2⁵⁶. The best-fit substitution models were inferred as above. The model parameters were estimated during the analysis. Gene partitions were unlinked. The substitution rates were allowed to vary across partitions with the *prset ratepr = variable* option. We conducted two independent runs of Metropolis-coupled Markov-chain Monte Carlo (MCMC). Each run consisted of four Markov chains, one cold and three incrementally heated, starting from a random tree. The chains were run for 1 × 10⁶ generations and sampled every 10² generations. The first 2.5 × 10³ sampled trees (25%) were discarded because they were likely to have been sampled during the burn-in period where the posterior probability did not reach the stationary distribution. Tracer ver. 1.6⁵⁷ was used to check that the log likelihood (ln *L*) scores had converged on a stationary distribution. The average standard deviation of split frequencies was 0.001871, indicating convergence of Metropolis-coupled MCMC (<0.01 recommended in the instructions of MrBayes ver. 3.1.2).

Chronological analyses and fossil calibration points. The chronological analyses were conducted with the Bayesian uncorrelated lognormal relaxed-clock model implemented in BEAST ver. 1.6.2^{57,58}. The input file was generated in BEAUti ver. 1.6.2⁵⁷. The best-fit substitution models for each gene partition were the same as the BI phylogenetic analysis. The Yule process of speciation was set as a tree prior. Multiple fossil calibration points were set^{59,60}. Our sampling of diverse familial and subfamilial lineages within Eulipotyphla permitted the use of multiple eulipotyphlan fossils for some families and subfamilies³⁴. Roca *et al.*²⁵ examined *S. cubanus* using only non-eulipotyphlan fossils as calibration points to estimate divergence times among eulipotyphlans, implying the possibility that the substitution rate in the eulipotyphlan lineages may not have been estimated accurately. This is

because evolutionary rates differ among mammalian orders³⁶, and the lineage-specific rate variations can strongly influence estimated divergence times^{31,36}. Therefore, using non-eulipotyphlan fossils in the estimation of divergence times among eulipotyphlans is not preferable. In this study, according to the multiple fossil information³⁴, we set 57–59, 37–49, 25–33 and 34–37 Ma as the TMRCA of the Erinaceidae-Soricidae, Erinaceidae, Soricidae and Talpinae-Scalopinae clades, respectively, as calibration points for each lineage, which are phylogenetically closer to the focal lineages than previously examined and reduce the effect of evolutionary rate variation. The minimum ages of the fossil time range (57, 37, 25 and 34, respectively) were set to the zero offset value of the lognormal distribution, assuming the minimum bound of each clade. The mean value was set to 1, 6, 4 and 1.5, respectively, where the mean took the mid-value of each fossil range (e.g., $6 = 43$ [mid-value of 37 and 49] - 37). The standard deviations of the lognormal distributions were adjusted so that the 95% quantile took each maximum age (59, 49, 33 and 37, respectively). Hereafter, this calibration is 'Assumption 1'. In addition, we used 76 Ma as the TMRCA of the Eulipotyphla clade, as estimated previously²⁵. In addition, 76 Ma was used as the mean for the normal prior distribution, setting the standard deviation so that the 95% credible interval ranged from 75 to 77 Ma. Hereafter, this calibration is 'Assumption 2'. In this study, we estimated the divergence times with three calibration strategies, [1] Assumption 1 only, [2] both Assumptions 1 and 2, and [3] Assumption 2 only. Strategy [1] tested the previous estimate of 76 Ma as the basal divergence time of Eulipotyphla without using 76 Ma as a calibration point. Strategy [3] assumed a similar time scale as Roca *et al.*²⁵ to test the influence of the difference between mtDNA²⁵ and nucDNA (this study) on the divergence times. Finally, using strategy [2], we used all available calibration standards. We conducted three independent MCMC runs of 1×10^7 generations in BEAST, where each parameter was sampled every 10^3 generations. The BEAST log file was assessed using Tracer 1.6 to confirm the parameter convergence to the stationary distribution and check the efficient sample size (ESS) for each parameter. The initial 25% of samples from each run were excluded as those from the burn-in period. Then, the post-burn-in samples from the three runs were combined. Most ESS values for the TMRCA parameters exceeded 200 (almost all >150), with a few exceptions of ESS <100 (Supplementary Table S2).

References

1. Woods, C. A. & Sergile, F. E. *Biogeography of the West Indies – patterns and perspectives*, 2nd edition. CRC Press, Florida, 582 pp (2001).
2. Ricklefs, R. & Bermingham, E. The West Indies as a laboratory of biogeography and evolution. *Phil. Trans. R. Soc. B.* **363**, 2393–2413 (2008).
3. Hedges, S. B. Vicariance and dispersal in Caribbean biogeography. *Herpetologica* **52**, 466–473 (1996a).
4. Hedges, S. B. Historical biogeography of West Indian vertebrates. *Annu. Rev. Ecol. Syst.* **27**, 163–196 (1996b).
5. Hedges, S. B. Biogeography of the West Indies: an overview. In (Woods, C. A. & Sergile, F. E. Eds) *Biogeography of the West Indies: patterns and perspectives*. CRC Press, Boca Raton, pp. 15–33 (2001).
6. Hedges, S. B. Paleogeography of the Antilles and origin of West Indian terrestrial vertebrates. *Ann. Missouri Bot. Gard.* **93**, 231–244 (2006).
7. Iturralde-Vinent, M. A. & MacPhee, R. D. E. Paleogeography of the Caribbean region: implications for Cenozoic biogeography. *Bull. Am. Mus. Nat. Hist.* **238**, 1–95 (1999).
8. MacPhee, R. D. E. & Iturralde, M. A. A short history of Greater Antillean land mammals: biogeography, paleogeography, radiations, and extinctions. *Tropics* **10**, 145–154 (2000).
9. Dávalos, L. M. Phylogeny and biogeography of Caribbean mammals. *Biol. J. Linn. Soc.* **81**, 373–394 (2004).
10. Ali, J. R. Colonizing the Caribbean: is the GAARlandia land-bridge hypothesis gaining a foothold? *J. Biogeogr.* **39**, 431–433 (2012).
11. Iturralde-Vinent, M. A. Meso-Cenozoic Caribbean paleogeography: implications for the historical biogeography of the region. *Int. Geol. Rev.* **48**, 791–827 (2006).
12. Crews, S. C. & Gillespie, R. G. Molecular systematics of *Selenops* spiders (Araneae: Selenopidae) from North and Central America: implications for Caribbean biogeography. *Biol. J. Linn. Soc.* **101**, 288–322 (2010).
13. Alonso, R., Crawford, A. J. & Bermingham, E. Molecular phylogeny of an endemic radiation of Cuban toads (Bufonidae: *Peltophryne*) based on mitochondrial and nuclear genes. *J. Biogeogr.* **39**, 434–451 (2012).
14. Řičan, O., Piálek, L., Zardoya, R., Doadrio, I. & Zrzavý, J. Biogeography of the Mesoamerican Cichlidae (Teleostei: Heroini): colonization through the GAARlandia land bridge and early diversification. *J. Biogeogr.* **40**, 579–593 (2013).
15. Matos-Maravi, P. *et al.* Causes of endemic radiation in the Caribbean: evidence from the historical biogeography and diversification of the butterfly genus *Calisto* (Nymphalidae: Satyrinae: Satyrini). *BMC Evol. Biol.* **14**, 199 (2014).
16. McHugh, A., Yablonsky, C., Binford, G. & Agnarsson, I. Molecular phylogenetics of Caribbean *Micrathena* (Araneae: Araneidae) suggests multiple colonization events and single island endemism. *Invertebr. Syst.* **28**, 337–349 (2014).
17. Hedges, S. B., Hass, C. A. & Maxon, L. R. Caribbean biogeography: molecular evidence for dispersal in West Indian terrestrial vertebrates. *Proc. Natl. Acad. Sci. USA* **89**, 1909–1913 (1992).
18. Hass, C. A., Maxson, L. R. & Hedges, S. B. Relationships and divergence times of West Indian amphibians and reptiles: insights from albumin immunology. (Woods, C. A. & Sergile, F. E. Eds) *Biogeography of the West Indies: patterns and perspectives*. CRC Press, Boca Raton, pp. 157–174 (2001).
19. Chakrabarty, P. Systematics and historical biogeography of Greater Antillean Cichlidae. *Mol. Phylogenet. Evol.* **39**, 619–627 (2006).
20. Ottenwalder, J. A. Systematics and Biogeography of the West Indian genus *Solenodon*. In (Woods, C. A. & Sergile, F. E. Eds) *Biogeography of the West Indies: patterns and perspectives*. CRC Press, Boca Raton, pp. 253–329 (2001).
21. Soy, J. & Mancina, C. A. *Solenodon cubanus*. The IUCN Red List of Threatened Species 2008: e.T20320A9185957. <http://dx.doi.org/10.2305/IUCN.UK.2008.RLTS.T20320A9185957.en>. Downloaded on 08 October 2015 (2008).
22. Silva-Taboada, G., Suárez Duque, W. & Díaz Franco, S. *Compendio de los Mamíferos Terrestres Autóctonos de Cuba vivientes y Extinguidos*. Museo Nacional de Historia Natural. Boloña Press, Habana, Cuba. 465 pp (2007).
23. Douady, C. J. & Douzery, E. J. P. Hedgehogs, shrews, moles, and solenodons (Eulipotyphla). In (Hedges, S. B. & Kumar, S. eds.) *The timetree of life*, Oxford University Press, pp. 495–498 (2009).
24. MacPhee, R. D. E. 'First' appearances in the Cenozoic land-mammal record of the Greater Antilles: significance and comparison with South American and Antarctic records. *J. Biogeogr.* **32**, 551–564 (2005).
25. Roca, A. L. *et al.* Mesozoic origin for west indian insectivores. *Nature* **429**, 649–651 (2004).
26. Steppan, S. J., Adkins, R. M., Spinks, P. Q. & Hale, C. Multigene phylogeny of the Old World mice, Murinae, reveals distinct geographic lineages and the declining utility of mitochondrial genes compared to nuclear genes. *Mol. Phylogenet. Evol.* **37**, 370–388 (2005).
27. Burrell, A. S., Disotell, T. R. & Bergey, C. M. The use of museum specimens with high-throughput DNA sequencers. *J. Hum. Evol.* **79**, 35–44 (2015).

28. Madsen, O. *et al.* Parallel adaptive radiations in two major clades of placental mammals. *Nature* **409**, 610–614 (2001).
29. Murphy, W. J. *et al.* Molecular phylogenetics and the origins of placental mammals. *Nature* **409**, 614–618 (2001).
30. Sato, J. J. *et al.* Deciphering and dating the red panda's ancestry and early adaptive radiation of Musteloidea. *Mol. Phylogenet. Evol.* **53**, 907–922 (2009).
31. Meredith, R. W. *et al.* Impacts of the Cretaceous Terrestrial Revolution and KPg extinction on mammal diversification. *Science* **334**, 521–524 (2011).
32. Bergsten, J. A review of long-branch attraction. *Cladistics* **21**, 163–193 (2005).
33. Rosen, D. E. A vicariance model of Caribbean biogeography. *Syst. Zool.* **24**, 431–464 (1975).
34. Douady, C. J. & Douzery, E. J. P. Molecular estimation of eulipotyphlan divergence times and the evolution of “Insectivora”. *Mol. Phylogenet. Evol.* **28**, 285–296 (2003).
35. Whidden, H. P. & Asher, R. J. The origin of the Greater Antillean Insectivorans. In (Woods, C. A. & Sergile, F. E. Eds) *Biogeography of the West Indies: patterns and perspectives*. CRC Press, Boca Raton, pp. 237–252 (2001).
36. Phillips, M. J. Geomolecular dating and the origin of placental mammals. *Syst. Biol.* **65**, 546–567 (2016).
37. Graham, A. Geohistory models and Cenozoic paleoenvironments of the Caribbean region. *Syst. Bot.* **28**, 378–386 (2003).
38. Brandt, A. H. *et al.* Mitogenomic sequences support a north-south subspecies subdivision within *Solenodon paradoxus*. Mitochondr. DNA doi: 10.3109/24701394.2016.1167891 (2016).
39. Arbogast, B. S., Edwards, S. V., Wakeley, J., Beerli, P. & Slowinski, J. B. Estimating divergence times from molecular data on phylogenetic and population genetic timescales. *Annu. Rev. Ecol. Syst.* **33**, 707–740 (2002).
40. Fabre, P.-H. *et al.* Rodents of the Caribbean: origin and diversification of hutias unraveled by next-generation museomics. *Biol. Lett.* **10**, doi: 10.1098/rsbl.2014.0266 (2014).
41. Santiago-Valentin, E. & Olmstead, R. G. Historical biogeography of Caribbean plants: introduction to current knowledge and possibilities from a phylogenetic perspective. *Taxon* **53**, 299–319 (2004).
42. Echelle, A. A., Fuselier, L., Van Den Bussche, R. A., Rodriguez, C. M. L. & Smith, M. L. Molecular systematics of Hispaniolan pupfishes (Cyprinodontidae: *Cyprinodon*): implications for the biogeography of insular Caribbean fishes. *Mol. Phylogenet. Evol.* **39**, 855–864 (2006).
43. Heinicke, M. P., Duellman, W. E. & Hedges, S. B. Major Caribbean and Central American frog faunas originated by ancient oceanic dispersal. *Proc. Natl. Acad. Sci. USA* **104**, 10092–10097 (2007).
44. Glor, R. E., Losos, J. B. & Larson, A. Out of Cuba: overwater dispersal and speciation among lizards in the *Anolis carolinensis* subgroup. *Mol. Ecol.* **14**, 2419–2432 (2005).
45. Hedges, S. B., Couloux, A. & Vidal, N. Molecular phylogeny, classification, and biogeography of West Indian racer snakes of the tribe Aphihiini (Squamata, Dipsadidae, Xenodontinae). *Zootaxa* **2067**, 1–28 (2009).
46. Brace, S., Turvey, S. T., Weksler, M., Hoogland, M. L. P. & Barnes, I. Unexpected evolutionary diversity in a recently extinct Caribbean mammal radiation. *Proc. R. Soc. B* **282**, 20142371, doi: 10.1098/rspb.2014.2371 (2015).
47. Dávalos, L. M. Short-faced bats (Phyllostomidae: Stenodermatina): a Caribbean radiation of strict frugivores. *J. Biogeogr.* **34**, 365–375 (2007).
48. Kay, R. F. Biogeography in deep time – What do phylogenetics, geology, and paleoclimate tell us about early platyrrhine evolution? *Mol. Phylogenet. Evol.* **82**, 358–374 (2015).
49. Oneal, E., Otte, D. & Knowles, L. L. Testing for biogeographic mechanisms promoting divergence in Caribbean cricket (genus *Amphiacusta*). *J. Biogeogr.* **37**, 530–540 (2010).
50. Forsman, A., Merilä, J. & Ebenhard, T. Phenotypic evolution of dispersal-enhancing traits in insular voles. *Proc. R. Soc. B* **278**, 225–232 (2011).
51. Tamura, K., Stecher, G., Peterson, D., Filipiński, A. & Kumar, S. MEGA6: Molecular Evolutionary Genetics Analysis Version 6.0. *Mol. Biol. Evol.* **30**, 2725–2729 (2013).
52. Wiens, J. J. Missing data and the design of phylogenetic analyses. *J. Biomed. Inform.* **39**, 34–42 (2006).
53. Wolsan, M. & Sato, J. J. Effects of data incompleteness on the relative performance of parsimony and Bayesian approaches in a supermatrix phylogenetic reconstruction of Mustelidae and Procyonidae (Carnivora). *Cladistics* **26**, 168–194 (2010).
54. Lanfear, R., Calcott, B., Ho, S. Y. W. & Guignon, S. PartitionFinder: Combined Selection of partitioning schemes and substitution models for phylogenetic analyses. *Molecular Biology and Evolution* **29**, 1695–1701 (2012).
55. Zwickl, D. J. Genetic Algorithm Approaches for the Phylogenetic Analysis of Large Biological Sequence Datasets under the Maximum Likelihood Criterion. PhD Thesis, University of Texas at Austin (2006).
56. Ronquist, F. & Huelsenbeck, J. P. MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* **19**, 1572–1574 (2003).
57. Drummond, A. J. & Rambaut, A. BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evol. Biol.* **7** (214), 1–8 (2007).
58. Drummond, A. J., Ho, S. Y. W., Phillips, M. J. & Rambaut, A. Relaxed phylogenetics and dating with confidence. *PLoS Biol.* **4**, 699–710 (2006).
59. Ho, S. Y. W. Calibrating molecular estimates of substitution rates and divergence times in birds. *J. Avian Biol.* **38**, 409–414 (2007).
60. Ho, S. Y. W. & Phillips, M. J. Accounting for calibration uncertainty in phylogenetic estimation of evolutionary divergence times. *Syst. Biol.* **58**, 367–380 (2009).

Acknowledgements

We are grateful to Antonio Cádiz Díaz, Emanuel Mora Macías, Lainet García Rivera, Manuel Iturriaga Monsibay, Kento Tanaka, and Yasunori Yamaguchi for their help throughout this study. We thank Osamu Hoson and Osamu Takahashi for their role in sampling *S. cubanus*. We also thank Robert C. Dowler for providing tissue samples for genetic analysis. This study was supported financially in part by JSPS via Grants-in-Aid for Scientific Research (21681028 and 25702047 to M.K.), the Naito Foundation (to M.K.), and Green Science Research Center in Fukuyama University (grants-in-aid for scientific research to J.J.S.). Other funding came from private contributions of the authors. Logistic support was provided by the Faculty of Biology, Havana University, Direction of Alejandro de Humboldt National Park at CITMA-Guantanamo, the Sector La Melba of Alejandro de Humboldt National Park, and the Center for Environmental Management and Control, The Environment Agency, Cuba.

Author Contributions

S.D.O. planned the research project. J.J.S. and M.K. financially supported the project. S.D.O., L.M.E.-D., R.B.-P., G.B.-Q., J.L.D.-L., J.G.-D., J.A.-L., S.T.N., N.Y. and M.K. did the sampling of seven Cuban solenodons. J.J.S. did the laboratory works, conducted the molecular phylogenetic analyses, and drafted the manuscript. All authors reviewed the manuscript and joined the discussion.

Additional Information

Supplementary information accompanies this paper at <http://www.nature.com/srep>

Competing financial interests: The authors declare no competing financial interests.

How to cite this article: Sato, J. J. *et al.* Molecular phylogenetic analysis of nuclear genes suggests a Cenozoic over-water dispersal origin for the Cuban solenodon. *Sci. Rep.* **6**, 31173; doi: 10.1038/srep31173 (2016).



This work is licensed under a Creative Commons Attribution 4.0 International License. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>

© The Author(s) 2016