

# Museum specimens shedding light on the evolutionary history and hidden diversity of the hedgehog family Erinaceidae

Ying Zeng

Kai He

Xing Chen

Weipeng Bai

Hongzhou Lin

Jianhai Chen

Nedko Nedyalkov

Nobuyuki Yamaguchi

Keerthy Vijayan

Ramamoorthy Suganthasakthivel

Brawin Kumar

Yuqing Han

Zhongzheng Chen

Wenzhi Wang

Yang Liu (✉ [liuy353@mail.sysu.edu.cn](mailto:liuy353@mail.sysu.edu.cn))

Sun Yat-sen University <https://orcid.org/0000-0003-4580-5518>

---

## Article

**Keywords:** divergence time, fossil species, mitogenome, museomics, phylogeny, small mammals

**Posted Date:** October 14th, 2022

**DOI:** <https://doi.org/10.21203/rs.3.rs-2160585/v1>

**License:**  This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

---

# Abstract

The family of Erinaceidae comprises 26 extant species in the subfamily Erinaceinae of spiny hedgehogs and the subfamily Galericinae of silky-furred gymnures and moonrats. These animals inhabit various habitats from tropical forests to deserts in Eurasia and Africa. Previous studies hinted that species diversity was likely underestimated. Moreover, erinaceids are among the oldest known living placental mammals originating more than 60 million years ago. The rich fossil records represent both living subfamilies and an extinct subfamily Brachyericinae. Comprehensive understanding of evolutionary history and taxonomic diversity is hampered by the unavailability of samples and the incorporation of molecular and morphological data. Here, we sequenced mitochondrial genomes from museum specimens and morphological data to reconstruct the genealogical relationships of Erinaceidae. Our results finely resolved interspecific relationships of living species and unveiled underestimated species diversity not only in *Hylomys* as revealed in previous studies, also in *Neotetracus* gymnures and *Atelerix*, *Hemiechinus*, and *Paraechinus* hedgehogs. The two extinct subfamilies, Brachyericinae and Erinaceinae were supported as sister taxa. There is a hint of a close relationship between fossil *Galerix* and Southeast Asian *Hylomys*. These findings highlight the potentiality of museomics but also found an overestimation of divergence times using mitogenomes as revealed in previous studies.

## Introduction

Species diversity is an important component of biodiversity and understanding global species pattern is a fundamental task for ecologists, evolutionary biologists, and conservationists<sup>1</sup>. Unbiased delineation of species diversity as well as their evolutionary history is essential for conservation planning and management<sup>1</sup>. Vicariance after initial colonization and resulting in allopatric speciation is one of the main species diversification processes<sup>2</sup>. Vicariance could be triggered by environmental changes due to climatic/geographic changes<sup>3</sup>. The divergence processes could not be well understood without a well-resolved phylogenetic tree<sup>4,5</sup>.

Eulipotyphla is an order of small-bodied insectivorous mammals (e.g., moles, hedgehogs, shrews and solenodons) with an evolutionary history of more than 85.8 million years<sup>6</sup>. Although more than 474 species have been recognized<sup>7</sup>, it is suggested that this order still comprises of a high proportion of hidden species<sup>8</sup>, which are unevenly distributed through the world and many of which are out of the existing protected areas<sup>9</sup>.

Erinaceidae thereof is one of the Eulipotyphla families that comprised morphologically distinct animals including the primitive furry forms moonrats and gymnures as well as spiny hedgehogs in two extant subfamilies, Galericinae and Erinaceinae. Currently, 26 species are recognized including eight Galericines and 18 hedgehogs. The evolutionary relationship is not entirely clear, and the taxonomic status of the recognized species is still under debate. For example, several species have never been sequenced or included in a molecular phylogenetic study, such as *Paraechinus nudiventris* and *Paraechinus microps*

from Southern Asia<sup>11,12</sup>, and the newly recognized *Mesechinus wangi* from southwestern China<sup>19</sup>, so examination of their taxonomic status and evolutionary history is warranted. Besides, previous studies have illustrated deep divergence in *Hylomys* indicating underestimated diversity<sup>10,13</sup>. However, the results relied on only partial fragments of the mitochondrial cytochrome b gene. It would be important to re-examine the pattern by capturing genome-wide signals across a broader taxon sampling.

The evolutionary relationship between fossils and living taxa is of great interest because of the high ratio between fossil and extant taxonomic diversity. While extant gymnures and moorhats are mainly distributed in South-eastern Asia, fossil records show Galericinae once flourished throughout Eurasia<sup>14</sup>. Species in the fossil subfamily Brachyericinae were once distributed in both North America and Eurasia. *Metechinus* and *Brachyerix* from the Miocene of North America were the earliest representatives<sup>15</sup>. This subfamily was completely extirpated in both continents around 5 million years ago<sup>15</sup>. Revealing the evolutionary relationships between living and fossil species could help to understand the dispersal history and factors shaping the mega extinction in Eurasia and North America.

Recently, there has been a boom in museomics<sup>16</sup>. Thanks to the developed Next-Generation Sequencing (NGS), which makes it possible to sequence highly fragmented DNA extracted from specimens in natural history collections and obtain a large amount of genomic information<sup>17</sup>. Several studies have explored the evolutionary history and addressed the taxonomic hypotheses, e.g., phylogenomics of the world's otters using modern and museum samples<sup>18</sup>. Due to the difficulty in obtaining overall living specimens of Erinaceidae, the museum collection provides an invaluable resource to investigate the evolutionary history of Erinaceidae.

In this study, we sequenced the complete mitochondrial genomes for 29 individuals representing 18 recognized species, 13 of which are historical specimens collected between 1909 and 1977 and deposited in the National Museum of Natural History (USNM). By combining with previously sequenced species, 23 out of 26 living species were included in the present analyses including divergence time estimation. We also examined the relationships between fossil and living species using a morphological data matrix. Taken together, we provide the most comprehensive phylogenetic work ever for living and fossil Erinaceids and unveil hidden species diversity in this speciose family.

## Results

### Mitogenome sequences de-contamination

We successfully determined complete mitogenomes (~ 98% completeness, except for the highly variable regions of the D-loop) for 29 samples. For destructive museum samples, we successfully sequenced 10 out of 16 samples (Table S4). Newly obtained sequences were submitted to GenBank (accession numbers XXXX-XXXXX).

We firstly estimated a mitogenome tree using RAxML and calculated Partitioned Bremer support per node to detect and mask potentially contaminated regions (Table S3). After de-contamination, the overall value of negative PBS increased (-263 → -171). Although several ingroup nodes were still characterized by non-positive and even strong negative PBS values (i.e., PBS < -5) (Fig S1), no potential recombination event could be determined. A sister relationship between *Neohylomys hainanensis* and *Neotetracus sinensis* was not strongly favored by any genes (i.e., PBS = 0; Fig S1), which may indicate a rapid diversification of the most recent common ancestor of *Hylomys*, *Neohylomys* and *Neotetracus*.

## Phylogenetic Relationships

The mitogenome RAxML tree covered 19 out of 26 recognized erinaceid species, three of which were sequenced for the first time including *Mesechinus wangi*, a species named in 2018 which was supported as in the *Mesechinus* clade and sister to *M. hughi* (Fig. 1A; BS = 100). *Paraechinus micropus* and *Paraechinus nudiventris* from South Asia were supported as sisters with each other (BS = 100) embedded within the *Paraechinus* clade. Importantly, all relationships at the species level or above were strongly supported (i.e., BS = 100). Fully resolved interspecific relationships in *Erinaceus*, *Mesechinus* and *Paraechinus*, provided a fine scaffold for supermatrix analyses.

Our supermatrix tree comprised 23 recognized species (Fig. 1B; missing three recognized species including *Podogymnura aureospinula* from the Philippines, *Atelerix sclateri* from Somalia, and *Hemiechinus collaris* from South Asia. The last species was sampled from the USNM but failed to be sequenced; Table S4). The supermatrix tree is incongruent with the mitogenome tree, and *Otohyalomys megalotis* (previously known as *Hylomys megalotis*; see Jenkin & Robinson, (2002)<sup>52</sup> is in a basal position of Galericinae, consistent with Bannivoka et al. (2014)<sup>19</sup>. Several relationships including a sister relationship between *Neohylomys* and *Neotetracus* were not highly supported, which is likely due to the inclusion of short sequences. In another hand, the inclusion of additional sequences elucidates deep intraspecific divergences in *Ne. sinensis*, *Hy. suillus*, *A. albiventris*, *P. aethiopicus* as well as in *H. auritus* (Fig. 3; Fig S2-S5).

To provide a complete snapshot of extant species diversity, we included the three missing species (see above) by integrating molecular and morphological matrices (see Method for detail). All three species grouped with their con-generic species (Fig. S6), supporting the morphological data provide phylogenetic information. We further included 11 fossil taxa to obtain an overview of the macroevolutionary history of erinaceids (Fig. 2). The fossil subfamily Brachyericinae was supported as a sister group to Erinaceinae with strong support. In Galericinae, *Neurogymnurus* and *Deinogalerix* were found as closely related to *Otohyalomys*, although this result is weakly supported. *Galerix* was supported as closely related to *Hylomys*. In Erinaceinae, *Ampechinus edwardsi* and *Gymnurechinus leakyi* were clustered together supporting the hypothesis that *Gymnurechinus* is closer to *Ampechinus* than other extant erinaceines<sup>53</sup>. *Mioechinus oeningensis* from the late Miocene (see Table S7) was closed related to *Paraechinus*, being

congruent with the previous study<sup>38</sup>. It is incongruent with the result of divergence time estimation supporting the MRCA of *Paraechinus* in the late Miocene.

## Divergence Times

We firstly estimated the divergence time using the complete mitogenomes (data set MtG). However, although we restricted the most recent common ancestor (MRCA) of erinaceids and shrews as in previous studies, the estimated MRCA of the two subfamilies were much older than in previous studies<sup>19,50</sup>. For example, the common ancestor of Erinaceinae was 23.08 Ma (95%CI = 13.82–33.57) as estimated using BEAST (Table S5). MCMCtree also provided a similar result (28.98Ma, 95%CI = 20.28–37.88) suggesting that the overestimation was not software sensitive. To test whether the result is due to saturation, we performed divergence time estimation using MtG<sup>2nd</sup>. This is because only the 2nd codon position did not show obvious saturation (Fig S16). The estimated MRCA of the two subfamilies were even older than using the complete MtG, suggesting the mitochondrial genes intrinsically tend to overestimate the divergence time as suggested in Zheng et al. (2011)<sup>51</sup>. Thirdly, we included 23 nuclear genes accompanying the MtG, to examine whether the slowly-evolving nuclear genes could “correct” the overestimation. The estimated MCRA of Erinaceinae were only 2–3 million years younger than using the MtG alone. Collectively, the overestimated divergence time of ingroup taxon is an intrinsic characteristic of mitogenome and could not be corrected using either “slowly evolving” codon positions or using extra slowly evolving nuclear genes. Finally, we included two additional calibrations to constrain the MCRAs of Erinaceinae and Galericinae, each. The estimated divergence times were eventually comparable with previous studies<sup>19,50</sup> (Fig. 4). For example, the common ancestor of Erinaceinae was 10.01 Ma (95%CI = 7.33–13.21) and 10.96 Ma (95%CI = 9.94–12.00) as estimated using BEAST and MCMCTree, respectively. In sum, the mitogenome overestimated the divergence time of ingroup species toward the calibration points, which might be because of saturation and could not be corrected using 2nd codon position or a combination of mitogenome and slow-evolving nuclear genes.

## Discussion

Applying the recent advanced DNA sequencing techniques to ancient and historic museum-preserved specimens and have greatly extended our understanding of biodiversity of organisms on earth and their evolutionary histories<sup>54</sup>, such as the mechanisms behind the evolutionary processes<sup>55,56</sup>, as well as the consequences of anthropogenic intervention during the Anthropocene<sup>57</sup>. Although we sequenced only the mitogenomes of Erinaceidae in this study, the same approach could be applied to the sequencing of the entire genomes which is much more informative. For example, sequencing ~ 120-years-old museum specimens of the Christmas Island rat recovered 95% of its genome regions<sup>58</sup>.

On the other hand, it is worth noting that potential contamination could be introduced during DNA extraction and DNA library preparation when handling museum samples, as we observed cross-sample contamination in our data set. Thus, scalable and automatic approaches to decontaminated sequences

are warranted<sup>59</sup> and processing the sample in an ancient DNA laboratory is a promising approach to control contamination<sup>60</sup>.

Although our results supported that mitogenomes provided good phylogenetic information, they also revealed biased estimated divergence times toward the calibration points. This is likely due to saturation and shorter coalescent time<sup>51,61</sup>, which affected the estimation of divergence time in fast-evolving lineages at a time scale of only ~ 10 million years<sup>61</sup>, and the effect could not be eliminated using RY-coded (AG→R; CT→Y) data or 2nd codon position data<sup>51,62</sup>. As such, ingroup calibrations and slow-evolving nuclear genes may be necessary to accurately estimate divergence time.

Our comprehensive analyses supported the three extant species without DNA sequences as in their congeneric clades (Fig S1), and *M. oeningensis* as closely related to *Paraechinus*, being consistent with a previous study<sup>38</sup>. These results suggesting the morphological data do provide phylogenetic information at least when there is a close relative of the fossil species. The sister relationship between Brachyericinae and Erinaceinae is incongruent with the original result of placing Brachyericini in Erinaceinae<sup>41,63</sup>. Note that without constraining the monophyly of the subfamily, *Dimylechinus bernoullii* was found in the Galericinae clade (Fig S14). This species was originally considered as closer to *Amphechinus edwardsi*<sup>43</sup>, and was placed under Brachyericinae<sup>64,41,42</sup>. Several novel relationships were found in Galericinae such as a close relationship between *Otohyalomys*, *Neurogymnurus* and *Deinogalerix*, and a close relationship between *Galerix* and *Hylomys*. *Galerix* is mainly distributed in Europe, and *Hylomys* are exclusively from Southeast Asia and Southern China. This relationship is only weakly supported [posterior probability (PP) = 62: Fig. 2] and contradictory to the hypothesis suggesting that *Galerix* and *Deinogalerix* are closely related<sup>65</sup>. However, together with the recent findings of *Galerix* spp. in South<sup>66,67</sup> and Southeast Asia<sup>68</sup>, the phylogenetic positions of *Neurogymnurus*, *Deinogalerix* as well as *Galerix* in Galericinae provides new clues of evolutionary relationships between living and fossil Galericines and biogeographic history of the subfamily.

Noteworthy, most fossil taxa do not have a close relative living species to align with; and the fossil specimens were commonly represented by fewer available (70–92 in our dataset; Table S9) and a large portion of missing characters. Indeed, the result obtained in the current study does not fully agree with the result estimated using maximum parsimony in the previous studies<sup>38</sup> suggesting the result is algorithm sensitive. Further, pseudo-extinction analyses found after removing molecular data of extant taxa from a morphological-molecular supermatrix, only 21%-42% of placental orders retained the same inter-ordinal placement, and this is especially true for Eulipotyphla<sup>69</sup>. In sum, although we tend to provide the phylogenetic positions of fossil species, these relationships need to be carefully re-examined.

Our results provide novel information regarding the diversification of extant species. Firstly, it supports that *P. micropus* and *P. nudiventris* are distinctive (Figs. 1). The two species inhabit western (i.e., Rajasthan Gujarat) and southern India, respectively<sup>11,12</sup>, and are separated by geographic barriers such as north of Cauvery, south of Godavari and south of the Mahanadhi rivers in the Eastern Ghats, and

Satpuras and the Chota Nagpur plateau (CN) in central India. The estimated divergence time between these two species (Fig. 4) in the Pleistocene is congruent with the separation time of Rajasthan-Gujarat in the Pleistocene.

The deep divergence of *Hy. suillus* started in the latest Miocene (Fig. 4), and has been repeatedly revealed<sup>10,13</sup>. All recognized subspecies of *Hy. suillus* as well as two unnamed lineages are likely distinct species (Fig. 3). We also revealed deep divergences in *Ne. sinensis* (Fig S4), *H. auritus* (Fig S2), *A. albiventris* (Fig S3), *P. aethiopicus* (Fig S5, also see O'Meara et al., 2021<sup>70</sup>) as well as the paraphyletic *M. hughi*, in all of which the distinct lineages are also allopatrically distributed indicating geographic isolation and underestimated species diversity. The two lineages of *M. hughi* are represented by specimens from northern China (Shaanxi province, near the type locality) and Southeastern China (represented by a sole sequence from a newly discovered distribution area of *M. hughi*, see Chen et al., (2020)<sup>71</sup>). Three lineages were found in *A. albiventris* (Fig S3), one was represented by a single sample from Senegal, westernmost Africa<sup>19</sup>, and the other two collected from western and eastern Africa were represented by specimens deposited in the USNM (Table S2). The mandible horizontal ramous of USNM573941 (Niger; skull broken) is slender and the two mental foramina are well separated below p1 and p4, while the horizontal ramous of USNM350004 (Sudan) and 325883 (Nigeria) are stouter and the two mental foramina are close with each other under p4 (fig S15). These characters have not been examined in previous studies. The taxonomy of these taxa needs to be revisited.

On the other hand, the *E. gymnura* from Indo-Malay Peninsula, and Boneo are genetically very closed (P-distance < 0.9%), even the populations are isolated by the Malacca and Karimata Straits, and have distinct pelage colours (dark brown vs. white). The result indicated the shallow Straits (no deeper than 50m) did not act as a barrier to *E. gymnura* dispersal and the leucism is a recent event.

The robust phylogeny combined with molecular dating also provides insight into the speciation due to migration. The divergence between *M. dauuricus* and the ancestor of *M. hughi*+*M. wangi* was at 2.39 Ma (95% CI = 1.29–3.69), and the MRCA of *M. hughi* and *M. wangi* was at 1.35Ma (95% CI = 0.62–2.21). This is congruent with the earliest record of *M. hughi* found in Early Pleistocene deposits in Tianzhen County, Shanxi Province<sup>72</sup>. The ancestors of *M. wangi* and *M. hughi* from Anhui (central east China<sup>71</sup>) are likely migrated southward due to the increased cooling and aridification in the middle Pleistocene (known as the middle Pleistocene transition, or MPT), dated to 1.2Ma-0.5Ma<sup>73,74</sup>. Interestingly, a fossil species *Mesechinus koloshanensis* from Geleshan, Chongqing<sup>72</sup> were overall similar to *M. hughi*, with a few characters assembling that of *M. wangi*. For instance, the P3 protocone of *M. koloshanensis* degenerates, but well-developed in *M. hughi*; and the paracristid on p4 of *M. koloshanensis* is shallower than that of *M. hughi*. These two characters are similar to those of *M. wangi*, indicating *M. koloshanensis* might represent an intermediate form between *M. hughi* and *M. wangi*. This finding suggested there were two or three corridors in China facilitating *Mesechinus* hedgehogs migrating southward and eventually resulting in speciation after colonization.

Collectively, our phylogenetic study indicates many hidden species and complex evolutionary history in Erinaceidae. It is especially important to reassess the population size, habitat area, and risk factors in the clade of *Neotetracus sinensis*, *Hylomys suillus*, *Atelerix albiventris*, *Paraechinus aethiopicus*, *Hemiechinus auratus*, particularly under the circumstance that hedgehogs are traded regularly throughout Africa, Eurasia, and South Asia and there are still no specific conservation measures for erinaceids<sup>75</sup>. Therefore, unveiling the hidden diversity in our study are a necessary first step for the decision on the conservation management.

## Materials And Methods

### Taxonomy and Taxon sampling

We followed Wilson and Mittermeier (2018)<sup>7</sup> for the taxonomy of extant erinaceids except for *Hylomys megalotis* which was recognized as in its own genus, namely *Otohylomys*<sup>19</sup> (Table S1). We sampled 29 tissues/specimens representing 18 out of the 26 recognized erinaceid species using museum destructive samples and newly collected fresh tissue samples (Table S2). Among these, *Mesechinus wangi*, *Paraechinus nudiventris* and *Paraechinus microps* were sequenced for the first time. A tissue sample of *Podogymnura truei* was loaned from the Field Museum of Natural History. Destructive sampling using museum specimens from the USNM was approved by the Department of Vertebrate Zoology; Division of Mammals (Invoice No. 2076134), National Museum of Natural history, Smithsonian Institution.

### Dna Extraction, Mitogenome Enrichment And Sequencing

The total DNA was extracted using 10% sodium dodecyl sulphate (SDS) with proteinase K, and purified using the phenol/chloroform approach<sup>76</sup>. For DNA extracted from fresh tissue samples, we sheared the DNA into small fragments using NEBNext dsDNA Fragmentise (New England Biolabs, Canada). This step was skipped for DNA extracted from museum samples. We constructed barcoded DNA libraries using a NEBNext Fast DNA Library Prep Set (New England Biolabs, Canada) with barcode adapters (NEXTflex DNA Barcodes, BIOO Scientific, USA). We purified the libraries using magnetic beads, size-selected using a 2% E-gel on an E-Gel electrophoresis system (Thermo Fisher Scientific, Canada), and re-amplified using a NEBNext High-Fidelity 2X PCR Master Mix (New England Biolabs, Canada).

We enriched mitochondrial genes using a capture-hybridization protocol following He et al. (2018)<sup>20</sup>. In brief, we first prepared biotin-linked DNA probes using long-range PCR amplicons. We hybridized our DNA libraries with probes at 65°C for 24–48 hours. We reamplified the enriched libraries before purification and concentration measurement using a Qubit Fluorometric Quantitation (Thermo Fisher Scientific, Canada). We ran the sequencing using a 316-chip on an Ion Torrent Personal Genome Machine (PGM). All laboratory work was conducted in China except for *P. nudiventris* which was conducted separately in India.



## Assembly, Annotation And Alignment

We trimmed the reads using Trimmomatic v0.36<sup>21</sup> to clean adaptor sequences at both ends, and thereafter corrected sequence error using BFC<sup>22</sup>. We assembled mitogenomes using SPAdes v3.11<sup>23</sup> and Geneious R11<sup>24</sup>, individually. For SPAdes assembly, we mapped the reads to a set of reference mitogenomes representing Erinaceidae species downloaded from GenBank using mrsFAST v3.3.0<sup>25</sup>. The mapped reads were assembled *de novo* using SPAdes with untrust contigs to reference as recommended by the manual. The assembled contigs were used as new references to which the reads were mapped for another run of mapping. We repeated the processes 3–20 times until all mitochondrial genes were recovered. We also assembled the sequences using Geneious R11 as described before<sup>13</sup>. In brief, the reads were mapped to reference genomes using the Geneious assembler iteratively up to 25 times. The results generated from the two assemblers were aligned using MAFFT and mismatch regions were checked carefully by eye. We annotated the mitogenomes using MITOS with default parameters<sup>26</sup>.

For mitogenome alignment, we first downloaded all available erinaceid mitogenomes from GenBank and selected *Solenodon paradoxus*, *Crocidura russula* and *Uropsilus gracilis* as outgroups due to their close relationships with *Erinaceidae* (Table S2). The two ribosomal RNA genes and 13 protein-coding genes were aligned using MAFFT v.7<sup>27</sup> as implemented in Geneious. The protein-coding genes (PCGs) were carefully checked by eye and manually corrected premature stop codons.

## Partitioned Bremer Support, Rdp5 And Saturation Analysis

Cross-contamination likely occurred during the DNA extraction and library preparation<sup>28</sup>. To minimize contaminated/unreliable sequences in our final alignment, we adopted a two-step strategy by calculating Partitioned Bremer supports (PBS) and detecting recombination using all 37 erinaceid mitogenomes and a corresponding mitogenome RAxML tree. Firstly, we calculated PBS for all genes on each node to detect the signal of recombination using PAUP v.4.0a168<sup>29</sup> and a Tcl script<sup>30</sup>. A positive PBS indicates that a given gene supports a particular node on the tree, and a negative PBS indicates a gene favours an alternative relationship<sup>31</sup>. When we observed a negative PBS < -5 for a given gene at any node of the tree, we conducted a recombination detection (RDP) using RDP V.5<sup>32</sup>. RDP detects whether a potential recombination breakpoint exists. Because of the maternal inheritance of the mammals' mitochondria, a “recombination” event suggests that one of the two parent fragments may be contaminated. RDP estimated UPGMA trees using the two fragments on both sides of a breakpoint (i.e., parent sequences), and we masked the parent sequences if the UPGMA tree violated our knowledge of the evolutionary history (Table S3). Finally, we estimated the PBS again using the masked alignment.

We examined the saturation of the coding genes, and for the 1st codon, 2nd codon and 3rd codon positions individually using DAMBE v.7.3.5<sup>33</sup> with an information entropy-based index of substitution saturation<sup>34</sup>. Then we plotted the transition versus transversion with F84 and GTR parameters.

# Molecular Phylogenies Using Mitogenome And Supermatrix Data

We conducted molecular phylogenetic analyses using i) the complete mitogenomes (including only the two rRNAs and 12 coding genes on the heavy chain, and excluded ND6 on the light chain), and ii) a gene supermatrix. The complete mitogenome data set (MtG hereafter) comprises 40 mitogenomes (including three outgroups). We further included mitochondrial genes (*cytochrome b*, *NADH dehydrogenase subunit 2* and 12S ribosomal RNA) sequences from GenBank to generate a sequence matrix using Geneious R11. This data set (supermatrix) comprises 132 erinaceid individuals.

We performed phylogenetic relationship analyses based on maximum-likelihood (ML) using RAxML-HPC v.8<sup>35</sup> and Bayesian inference (BI) using MrBayes v.3.2.7a<sup>36</sup>. The best partitioning scheme and associated models were estimated using PartitionFinder v.2.1.1<sup>37</sup> for each data set for RAxML and MrBayes individually. The results are given in Table S8. RAxML analyses were performed on XSEDE via the CIPRES web portal (<http://www.phylo.org>). *Solenodon paradoxus* was used to root the tree. We selected GTR + Gamma model in both trees search and bootstrapping phases. We didn't use the BFGS searching algorithm. We ran a rapid bootstrapping (-x) with recommended parameter 'autoMRE' to let the program halt bootstrapping automatically and search for the best-scoring ML tree (-f a). We ran MrBayes using four simultaneous Markov chain Monte Carlo (MCMC) chains including one cold chain and three hot chains, a temperature of 0.065, and runs of 5 million generations. Trees were sampled every 5000 generations and the burn-in fraction was set as 30%. We repeated the analyses twice to ensure convergence on the same states.

## Integrative Analyses On Dna And Morphological Data

We further used morphology-gene integrated data to estimate phylogenetic positions for living and fossil species that have no available genetic information. A morphological data matrix was generated by Gould<sup>38,39,40</sup> representing 19 living and 22 fossil species (19 living and 11 fossil species used in this study), and morphological data for *M. hughi* was included in He et al. (2012)<sup>13</sup>. We selected a subset of 53 individuals representing all distinct lineages and integrated it with morphological data for 26 living species (supplementary alignment1) to estimate the phylogenetic positions of three extant species (i.e., *Atalerix sclateri*, *Hemiechinus collaris*, *Podogymnura aureospinula*) that has no available DNA sequences. We performed Bayesian analyses in MrBayes. The partitioning strategy and other parameters were set as above.

We further included 11 fossil taxa provided by Gould, (1995)<sup>38</sup> (supplementary alignment2). The other 11 fossil taxa were not included because they were represented by too few characters (i.e., characters < 25), and because removing these "rogue taxa" could improve phylogenetic accuracy<sup>44</sup>. We ran MrBayes as mentioned above but constrained the following topologies: the monophyly of Erinaceidae, Galericinae, Brachyericinae, Erinaceinae as well as the monophyly of *Deinogalerix*, *Galerix* and *Gymnurechinus*. A brief taxonomic revision and systematic assignments for each fossil was fully explained in Table S7.

## Divergence Time Estimation

We estimated divergence times using the MCMCTree program in PAML 4.9i<sup>45</sup> and BEAST v.2.6.6<sup>46</sup>. For BEAST analyses, we partitioned the data based on the result of PartitionFinder as mentioned above (Table S8). The clock model and time tree were linked across partitions and a birth-death model was used for the tree prior. We performed MCMC runs of 50 million generations and the trees were sampled every 5000 generations. The two calibrations were set as followed: we used a lognormal prior distribution (mean 3.61, SD 0.142, offset 0; mean 4.27, SD 0.084, offset 0 respectively), rendering the time distribution covering the fossil age. We used Tracer v.1.7 to estimate the posterior distribution of each parameter in the log file to ensure that analyses reached stationary states. Then we used TreeAnnotator (BEAST package) to summarize the sampled trees into a maximum clade credibility tree with mean heights parameters. Analyses were conducted twice.

For MCMCTree analyses, we adjusted the time unit to 100 Ma. The ML tree topology with two calibrations prior was used as an input tree file (supplement timetree1 file). Then MCMCTree calculated the tree's branch length, the gradient vector and the Hessian under the model GTR + G (model = 7). We used a lognormal clock model and set the mean substitution rate (rgene\_gamma) and the rate drift parameter (sigma2\_gamma) as G (2 2000) and G (1, 4.5) respectively. The birth-death parameters were set as 0.01, 0.01, and 0.1. Each run discarded the first 50000 generations as burn-in and sampled every 50 iterations until it reached 5 million. Similarly, we used Tracer v.1.7 to examine the stationary state and convergence of each run. Analyses were conducted twice.

We firstly used two fossil-based calibrations. One is based on the *Adunator* from the Torrejonian between 61.1 Ma and 84.2 Ma<sup>6</sup>. This is the oldest known fossil of erinaceomorph and was considered as on the stem Erinaceomorpha<sup>47</sup>. The second calibration was according to fossils *Neurogymnurus indricotherii* and *Palaeosaptor acridens* both from the early Oligocene<sup>48, 49</sup>. *N. indricotherii* was the earliest unquestioned *Galericine* from western Kazakhstan in Asia. *P. acridens* from central Mongolia was devoted to the subfamily *Erinaceinae*<sup>49</sup>. We set the most recent common ancestor (MRCA) of the two subfamilies to 28.3 Ma – 48.8Ma referring to Meredith et al., (2011)<sup>6</sup>.

First, we ran BEAST and MCMCTree using the MtG data set. The estimated MRCA of both subfamilies were strikingly older than that estimated in previous studies<sup>13, 19</sup> (see Result and Table S5). We further tested whether it was because of substitution saturation by conducting substitution saturation plots using DAMBE v7 for the 12 coding genes as well as for each codon position (Fig S16). We extracted the 2nd codon position which was not saturated (data set MtG<sup>2nd</sup>) following Zheng et al. (2011)<sup>51</sup>. The estimated divergence times using MtG<sup>2nd</sup> were not obviously different from that using MtG, we further included available nuclear genes from He et al., (2021)<sup>50</sup> for erinaceids and outgroups. These included 17 nuclear gene fragments (~ 39443 bp) for six erinaceids and three outgroup taxa. Because this analysis still resulted in similar estimations for divergence times, we included two additional secondary calibrations on the MRCA of Erinaceinae and Galericinae in He et al. (2021)<sup>50</sup>. While using BEAST we

added a MRCA prior with a normal distribution (mean 17.71, sigma 2.05, offset 0; mean 6.97, sigma 2.05, offset 0 respectively). And using MCMCTree we added two extra calibration priors to the input tree file (supplement timetree2 file).

## Declarations

### Acknowledgement

We thank the curators and collection staff at the Field Museum of Natural history, and the National Museum of Natural History, Smithsonian Institution, for accessing specimens under their care and for permitting tissue sampling and destructive sampling. Any use of trade, product, or firm names is for descriptive purposes only and does not imply endorsement by the US government.

### Funding

This work was supported by the National Natural Science Foundation of China (31970389 and 32170452 to KH,31572251 to YL), Guangdong Natural Science Funds for Distinguished Young Scholar (2022B1515020033 to KH), and Forestry Administration of Guangdong Province, China (DFGP Project of Fauna of Guangdong-202115 to KH and YL).

### Author's contributions

KH and YL designed this study. KH and WW organized the molecular work, KH, JC, NN, NY, KV, RS, BK, ZC involved in fieldwork and contributed to sample collection and identification, XC and RS conducted molecular experiments, ZY, KH and XC, and HL performed data analyses, YZ, KH, XC, WPB and LY draft the MS, YH drew sampling figures.

### Conflict of interest

None declared.

### Data availability declared

The authors confirm that the data supporting the findings of this study are available within the article and its supplementary materials.

## References

1. Coates, D.J., Byrne, M., Moritz, C.: Genetic diversity and conservation units: Dealing with the species-population continuum in the age of genomics. *Front. Ecol. Evol.* **6**, (2018)
2. Gizaw, A., et al.: Vicariance, dispersal, and hybridization in a naturally fragmented system: the afro-alpine endemics *Carex monostachya* and *C. runssoroensis* (Cyperaceae). *Alp. Bot.* **126**, 59–71 (2016)

3. Hewitt, G.: The genetic legacy of the Quaternary ice ages. *Nat. Rev.* **405**, 907–913 (2000)
4. Faircloth, B.C., Sorenson, L., Santini, F., Alfaro, M.E.: A Phylogenomic Perspective on the Radiation of Ray-Finned Fishes Based upon Targeted Sequencing of Ultraconserved Elements (UCEs). *PLoS One* **8**, (2013)
5. Johnson, B.R., et al.: Phylogenomics Resolves Evolutionary Relationships among Ants, Bees, and Wasps. *Curr. Biol.* **23**, 2058–2062 (2013)
6. Meredith, R.W., et al.: Impacts of the cretaceous terrestrial revolution and KPg extinction on mammal diversification. *Sci. (80-)*. **334**, 521–524 (2011)
7. Wilson, D., Mittermeier, R.: Handbook of the mammals of the world: vol. 8: insectivores, sloths and colugos, IUCN: International Union for Conservation of Nature. in *IUCN, Conservation International, IUCN Species Survival Commission (SSC)* doi: (2018).  
10.1073/pnas.2120307119/-/DCSupplemental.Published
8. Parsons, D.J., Pelletier, T.A., Wieringa, J.G., Duckett, D.J., Carstens, B.C.: Analysis of biodiversity data suggests that mammal species are hidden in predictable places. *Proc. Natl. Acad. Sci. U. S. A.* **119**, (2022)
9. Kennerley, R.J., et al.: Global patterns of extinction risk and conservation needs for Rodentia and Eulipotyphla. *Divers. Distrib.* **27**, 1792–1806 (2021)
10. Ruedi, M., Fumagalli, L.: Genetic structure of Gymnures (genus *Hylomys*; Erinaceidae) on continental islands of Southeast Asia: historical effects of fragmentation. *J. Zoo Syst. Evol. Res.* **34**, 153–162 (1996)
11. Kumar, B., Togo, J.: Ranjit Singh. The South Indian hedgehog *Paraechinus nudiventris* (Horsfield, 1851): review of distribution data, additional localities and comments on habitat and conservation. *Mammalia*. **83**, 399–409 (2019)
12. Kumar, B., Togo, J., Singh, R.: Predicting the potential distribution of the lesser-known endemic Madras hedgehog *Paraechinus nudiventris* (Order: Eulipotyphla, Family: Erinaceidae) in southern India. *Mammalia*. **83**, 470–478 (2019)
13. He, K., et al.: An estimation of erinaceidae phylogeny: A combined analysis approach. *PLoS One* **7**, (2012)
14. Corbet, G.B.: The family Erinaceidae: a synthesis of its taxonomy, phylogeny, ecology and zoogeography. *Mamm. Rev.* **18**, 117–172 (1988)
15. Rich, T.H.V.: Origin and History of the Erinaceinae and Brachyericinae (Mammalia, Insectivora) in North America. *Bull. Am. MUSEUM Nat. Hist.* **171**, (1981)
16. Nattier, R.: Biodiversity in natural history collections: A source of data for the study of evolution. *Biodivers. Evol.* 175–187 (2018). doi:10.1016/B978-1-78548-277-9.50010-3
17. Liu, Y., Bennett, E.A., Fu, Q.: Evolving ancient DNA techniques and the future of human history. *Cell*. **185**, 2632–2635 (2022)
18. Ferran, V., De, et al.: Phylogenomics of the world's otters. *Curr. Biol.* **32**, 3650–3658 (2022)

19. Bannikova, A.A., Lebedev, V.S., Abramov, A.V., Rozhnov, V.V.: Contrasting evolutionary history of hedgehogs and gymnures (Mammalia: Erinaceomorpha) as inferred from a multigene study. *Biol. J. Linn. Soc.* **112**, 499–519 (2014)
20. He, K., et al.: A new genus of Asiatic short-tailed shrew (Soricidae, Eulipotyphla) based on molecular and morphological comparisons. *Zool. Res.* **39**, 309–323 (2018)
21. Bolger, A.M., Lohse, M., Usadel, B., Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics.* **30**, 2114–2120 (2014)
22. Li, H.B.F.C.: Correcting Illumina sequencing errors. *Bioinformatics.* **31**, 2885–2887 (2015)
23. Bankevich, A., et al.: SPAdes: A new genome assembly algorithm and its applications to single-cell sequencing. *J. Comput. Biol.* **19**, 455–477 (2012)
24. Olsen, C., Qaadri, K.: Geneious R7: A Bioinformatics Platform for Biologists. *Plant. Anim. Genome XXII Conf.* **22**, 9490 (2014)
25. Hach, F., et al.: MrsFAST-Ultra: A compact, SNP-aware mapper for high performance sequencing applications. *Nucleic Acids Res.* **42**, 494–500 (2014)
26. Bernt, M., et al.: Improved de novo metazoan mitochondrial genome annotation. *Mol. Phylogenet Evol.* **69**. **MITOS**, 313–319 (2013)
27. Katoh, K., Standley, D.M.: MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. *Mol. Biol. Evol.* **30**, 772–780 (2013)
28. Castañeda-Rico, S., León-Paniagua, L., Edwards, C.W., Maldonado, J.E.: Ancient DNA From Museum Specimens and Next Generation Sequencing Help Resolve the Controversial Evolutionary History of the Critically Endangered Puebla Deer Mouse. *Front. Ecol. Evol.* **8**, 1–18 (2020)
29. Swofford, D.L., Sullivan, J.: Phylogeny inference based on parsimony and other methods using PAUP. *Phylogenetic Handb.* 267–312 (2012). doi:10.1017/cbo9780511819049.010
30. Göker, M., Voglmayr, H., Blázquez, G.G., Oberwinkler, F.: Species delimitation in downy mildews: the case of *Hyaloperonospora* in the light of nuclear ribosomal ITS and LSU sequences. *Mycol. Res.* **113**, 308–325 (2009)
31. Lambkin, C.L., Lee, M.S.Y., Winterton, S.L., Yeates, D.K.: Partitioned Bremer support and multiple trees. *Cladistics.* **18**, 436–444 (2002)
32. Martin, D.P., et al.: RDP5: A computer program for analyzing recombination in, and removing signals of recombination from, nucleotide sequence datasets. *Virus Evol.* **7**, (2021)
33. Xia, X.: DAMBE6: New Tools for Microbial Genomics, Phylogenetics, and Molecular Evolution. *J. Hered.* 431–437 (2017). doi:10.1093/jhered/esx033
34. Xia, X., Xie, Z., Salemi, M., Chen, L., Wang, Y.: An index of substitution saturation and its application. *Mol. Phylogenet Evol.* **26**, 1–7 (2003)
35. Stamatakis, A.: RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics.* **30**, 1312–1313 (2014)

36. Ronquist, F., et al.: Mrbayes 3.2: Efficient bayesian phylogenetic inference and model choice across a large model space. *Syst. Biol.* **61**, 539–542 (2012)
37. Lanfear, R., Frandsen, P.B., Wright, A.M., Senfeld, T., Calcott, B.: Partitionfinder 2: New methods for selecting partitioned models of evolution for molecular and morphological phylogenetic analyses. *Mol. Biol. Evol.* **34**, 772–773 (2016)
38. Gould, G.C.: Hedgehog Phylogeny (Mammalia, Erinaceidae) - the Reciprocal Illumination of the Quick and the Dead. *Am. Museum Novit.* 1–45(1995)
39. Gould, G.C.: *Systematic revision of the Erinaceidae (Mammalia)-a comprehensive phylogeny based on the morphology of all known taxa. Columbia University, New York* doi:(1997). 10.1038/nrg3174
40. Gould, G.C.: The phylogenetic resolving power of discrete dental morphology among extant hedgehogs and the implications for their fossil record. *Am. Mus. Novit.* **3340**, 1–52 (2001)
41. Gureev, A.A., Insectivores (Mammalia, I.): Hedgehogs, Moles, and Shrews (Erinaceidae, Talpidae, Soricidae), *Fauna SSSR.Mlekopitayushchie (Fauna USSR Mammals)* **4**, (1979)
42. Li, C.K., Qiu, Z.D., Tong, Y.S., Zheng, S.H., Ni, X.J.: *Palaeovertebrata Sinica. in Science Press, Beijing* vol. 3 Fascicle (2015)
43. Hürzeler, J.: Beiträge zur Kenntnis der Dimylidae. *Schweiz. Paläont Abh.* **65**, 1–44 (1944)
44. Aberer, A.J., Krompass, D., Stamatakis, A.: Pruning Rogue Taxa Improves Phylogenetic Accuracy: An Efficient Algorithm and Webservice. *Syst. Biol.* **62**, 162–166 (2013)
45. Yang, Z.: PAML 4: Phylogenetic analysis by maximum likelihood. *Mol. Biol. Evol.* **24**, 1586–1591 (2007)
46. Bouckaert, R., et al.: BEAST 2.5: An advanced software platform for Bayesian evolutionary analysis. *PLoS Comput. Biol.* **15**, (2019)
47. Benton, M.J., Donoghue, P.C.J., Asher, R.J.: Calibrating and constraining molecular clocks. in *The Timetree of Life* 35–86 (2009)
48. Lopatin, A.V.: Oligocene and early miocene insectivores (Mammalia) from Western Kazakhstan. *Paleontol. J.* **33**, 182–191 (1999)
49. Ziegler, R., Stuttgart: The nyctitheriids (Lipotyphla, Mammalia) from Early Oligocene fissure fillings in South Germany. *Neues Jahrb. fur Geol. und Palaontologie - Abhandlungen.* **246**, 183–203 (2007)
50. He, K., et al.: Myoglobin primary structure reveals multiple convergent transitions to semi-aquatic life in the world's smallest mammalian divers. *Elife.* **10**, 1–27 (2021)
51. Zheng, Y., Peng, R., Kuro-O, M., Zeng, X.: Exploring patterns and extent of bias in estimating divergence time from mitochondrial DNA sequence data in a particular lineage: A case study of salamanders (Order Caudata). *Mol. Biol. Evol.* **28**, 2521–2535 (2011)
52. Jenkin, P.D., Robinson, M.F.: Another variation on the gymnure theme: description of a new species of *Hylomys* (Lipotyphla, Erinaceidae, Galericinae). *Bull. Nat. Hist. Museum Zool.* **68**, 1–11 (2002)
53. Butler, P.M.: The skull of ictops and the classification of the insectivora. *Proc. Zool. Soc. London* **126**, 453–481 (1956)

54. Raxworthy, C.J., Smith, B.T.: Mining museums for historical DNA: advances and challenges in museomics. *Trends Ecol. Evol.* **36**, 1049–1060 (2021)
55. Hahn, E.E., Grealy, A., Alexander, M., Holleley, C.E.: Museum Epigenomics: Charting the Future by Unlocking the Past. *Trends Ecol. Evol.* **35**, 295–300 (2019)
56. Rubi, T.L., Knowles, L.L., Dantzer, B.: Museum epigenomics: Characterizing cytosine methylation in historic museum specimens. *Mol. Ecol.* **20**, 1161–1170 (2020)
57. Schmitt, C.J., Cook, J.A., Zamudio, K.R., Edwards, S.V.: Museum specimens of terrestrial vertebrates are sensitive indicators of environmental change in the Anthropocene. *Philos. Trans. R. Soc. B Biol. Sci.* **374**, (2019)
58. Lin, J., et al.: Probing the genomic limits of de-extinction in the Christmas Island rat. *Curr. Biol.* **32**, 1650–1656 (2022)
59. Simion, P., et al.: A software tool ‘CroCo’ detects pervasive cross-species contamination in next generation sequencing data. *BMC Biol.* **16**, (2018)
60. Llamas, B., et al.: From the field to the laboratory: Controlling DNA contamination in human ancient DNA research in the high-throughput sequencing era. *STAR. Sci. Technol. Archaeol. Res.* **3:1**, 1–14 (2017)
61. Nilsson, M.A., Härlid, A., Kullberg, M., Janke, A.: The impact of fossil calibrations, codon positions and relaxed clocks on the divergence time estimates of the native Australian rodents (Conilurini). *Gene.* **455**, 22–31 (2010)
62. Phillips, M.J.: Branch-length estimation bias misleads molecular dating for a vertebrate mitochondrial phylogeny. *Gene.* **441**, 132–140 (2009)
63. Van Valen, L.: New paleocene insectivores and insectivore classification. *Bull. Am. Museum Nat. Hist.* **135**, 217–284 (1967)
64. Mckenna, M.C., Holton, C.P.: A New Insectivore from the Oligocene of Mongolia and a New Subfamily of Hedgehogs. *Am. Museum Nat. Hist.* 1–11(1967)
65. Borrani, A., Savorelli, A., Masini, F., Mazza, P.P.: A. The tangled cases of *Deinogalerix* (Late Miocene endemic erinaceid of Gargano) and *Galericini* (Eulipotyphla, Erinaceidae): a cladistic perspective. *Cladistics.* **34**, 542–561 (2018)
66. Zijlstra, J., Flynn, L.J., Hedgehogs: (Erinaceidae, Lipotyphla) from the Miocene of Pakistan, with description of a new species of *Galerix*. *Palaeobio Palaeoenv.* **95**, 477–495 (2015)
67. Parmar, V., Norboo, R., Magotra, R., Parmar, V.: First record of Erinaceidae and Talpidae from the Miocene Siwalik deposits of India First record of Erinaceidae and Talpidae from the Miocene Siwalik deposits of India. *Hist. Biol.* **00**, 1–8 (2022)
68. Cailleux, F., Chaimanee, Y., Jaeger, J., Chavasseau, O.: New Erinaceidae (Eulipotyphla, Mammalia) from the Middle Miocene of Mae Moh, Northern Thailand. *J. Vertebr. Paleontol.* **40**, (2020)
69. Brady, P.L., Springer, M.S.: The effects of fossil taxa, hypothetical predicted ancestors, and a molecular scaffold on pseudoextinction analyses of extant placental orders. *PLoS One* **16**, (2021)



70. O'Meara, D., O'Reilly, C., Abdullahi, A.A., Baker, M.A., Yamaguchi, N.: Phylogeography of desert hedgehogs (*Paraechinus aethiopicus*) in Qatar: Implications for its intra-specific phylogeny and taxonomy. *J. Arid Environ.* **J.193**, (2021)
71. Chen, Z., et al.: First record of genus *Mesechinus* (Mammalia: Erinaceidae) in Anhui Province, China- *Mesechinus hughii*. *Acta Theriologica Sin.* **20**, 96–99 (2020)
72. Bai, W., et al.: Pleistocene Hedgehog *Mesechinus* (Eulipotyphla, Mammalia) in China. *J. Mamm. Evol.* (2022). doi:10.1007/s10914-022-09612-w
73. Martin, R.A., et al.: Rodent community change at the Pliocene-Pleistocene transition in southwestern Kansas and identification of the *Microtus* immigration event on the Central Great Plains. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **267**, 196–207 (2008)
74. Clark, P.U., et al.: The middle Pleistocene transition: characteristics, mechanisms, and implications for long-term changes in atmospheric pCO<sub>2</sub>. *Quat Sci. Rev.* **25**, 3150–3184 (2006)
75. Nijman, V., Bergin, D.: Trade in hedgehogs (Mammalia: Erinaceidae) in Morocco, with an overview of their trade for medicinal purposes throughout Africa and Eurasia. *J. Threat Taxa.* **7**, 7131–7137 (2015)
76. Sambrook, J., Russell, D.: *Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory Press.* Elsevier Inc. (2001). doi:10.1016/j.yempv.2022.107620

## Figures

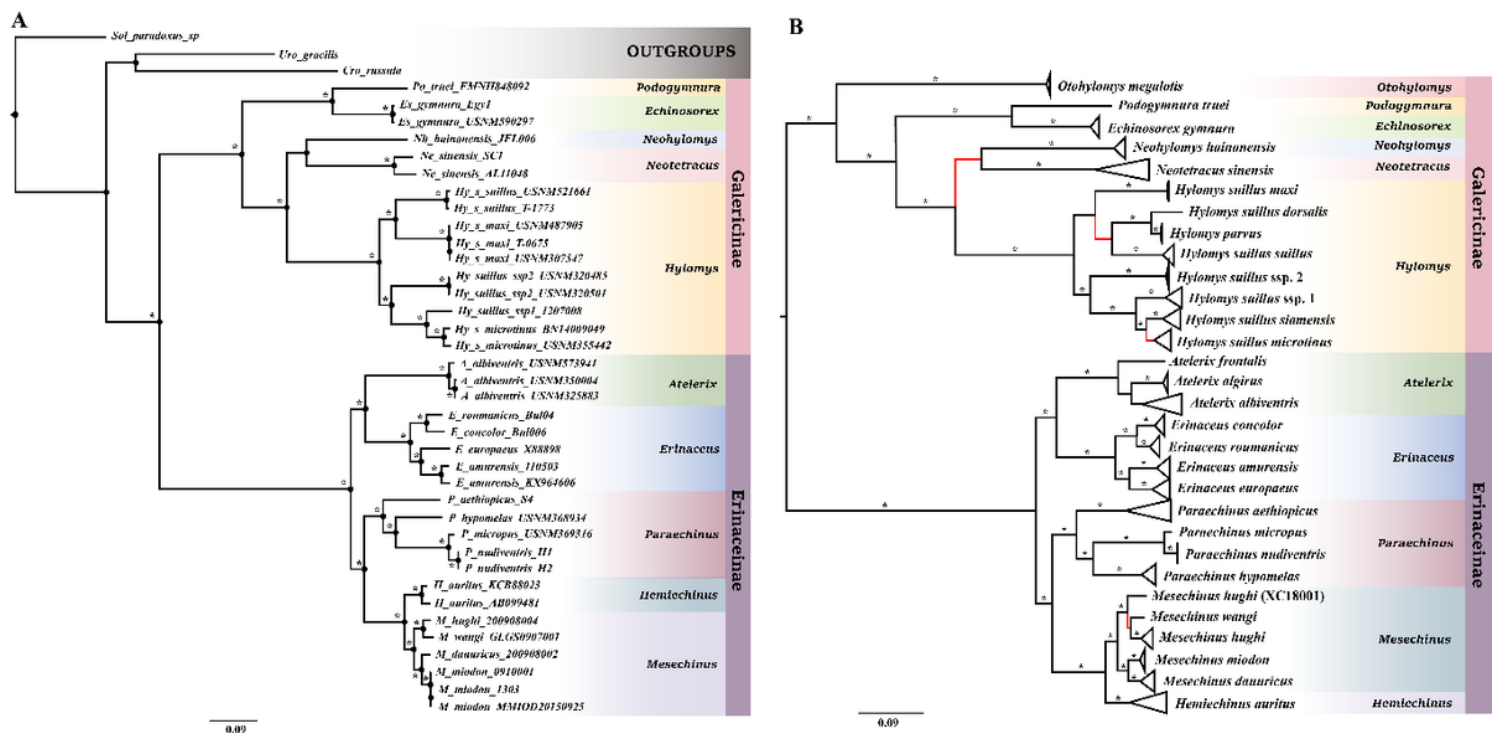


Figure 1

Maximum likelihood tree of Erinaceidae based on MtG dataset (A) and supermatrix dataset (B). In the Fig. 1A, a node with an asterisk indicates bootstrap proportion of the node is higher than or equal to 95. The remaining nodes that didn't label indicate the bootstrap proportion is between 70 and 95. In the Fig. 1B, whole numbers represent the bootstrap proportion of the branch. A branch with an asterisk indicates the bootstrap proportion of the branch is higher than or equal to 95. Red branch indicates the bootstrap proportion of the branch is lower than 70. The remaining branches that didn't label indicate the bootstrap proportions is between 70 and 95. Outgroups were omitted and multiple samplings of the same species or subspecies were collapsed into triangles. The name in the parentheses is an individual of the clade, which did not cluster with other individuals of the same species for some reason.

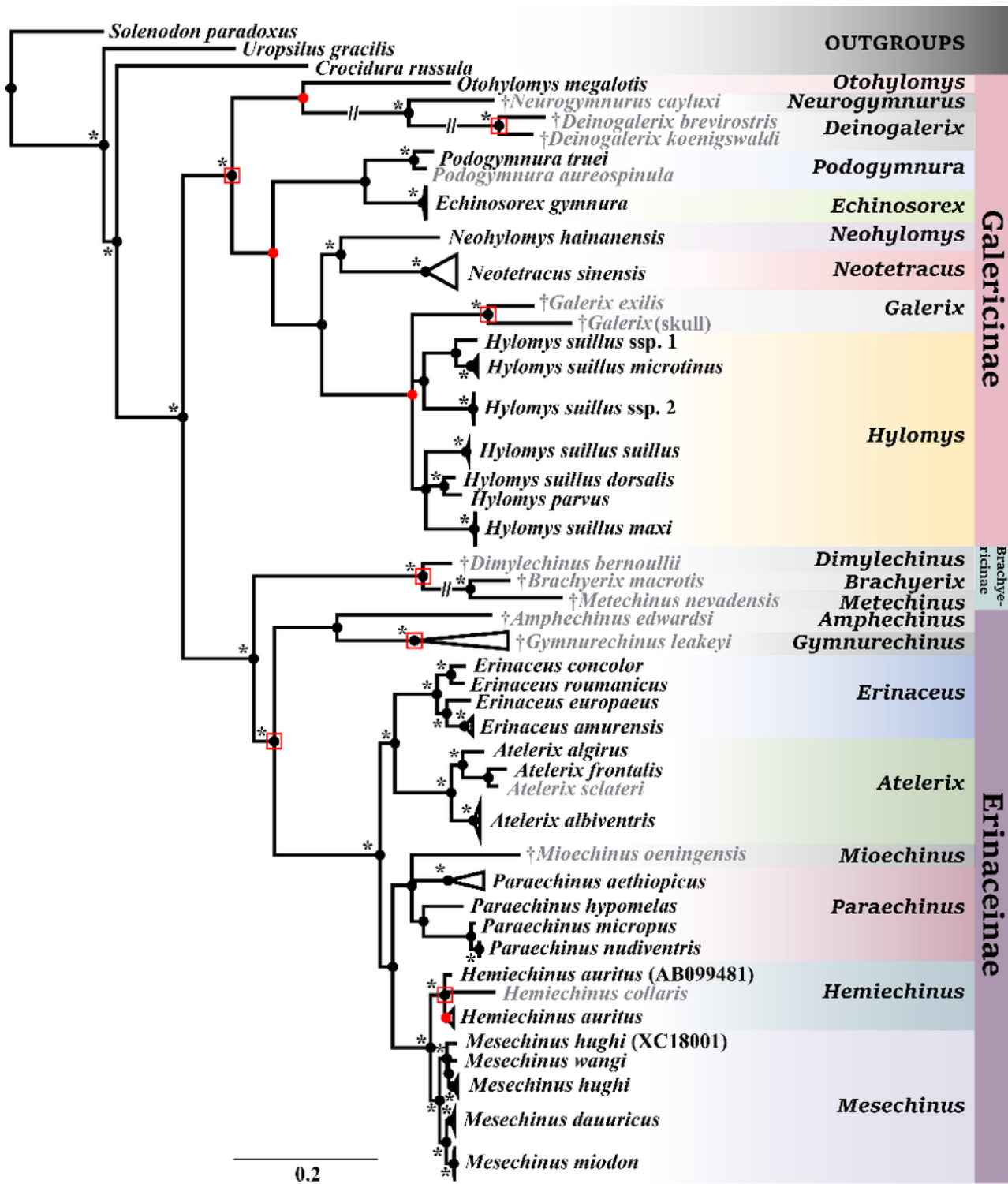
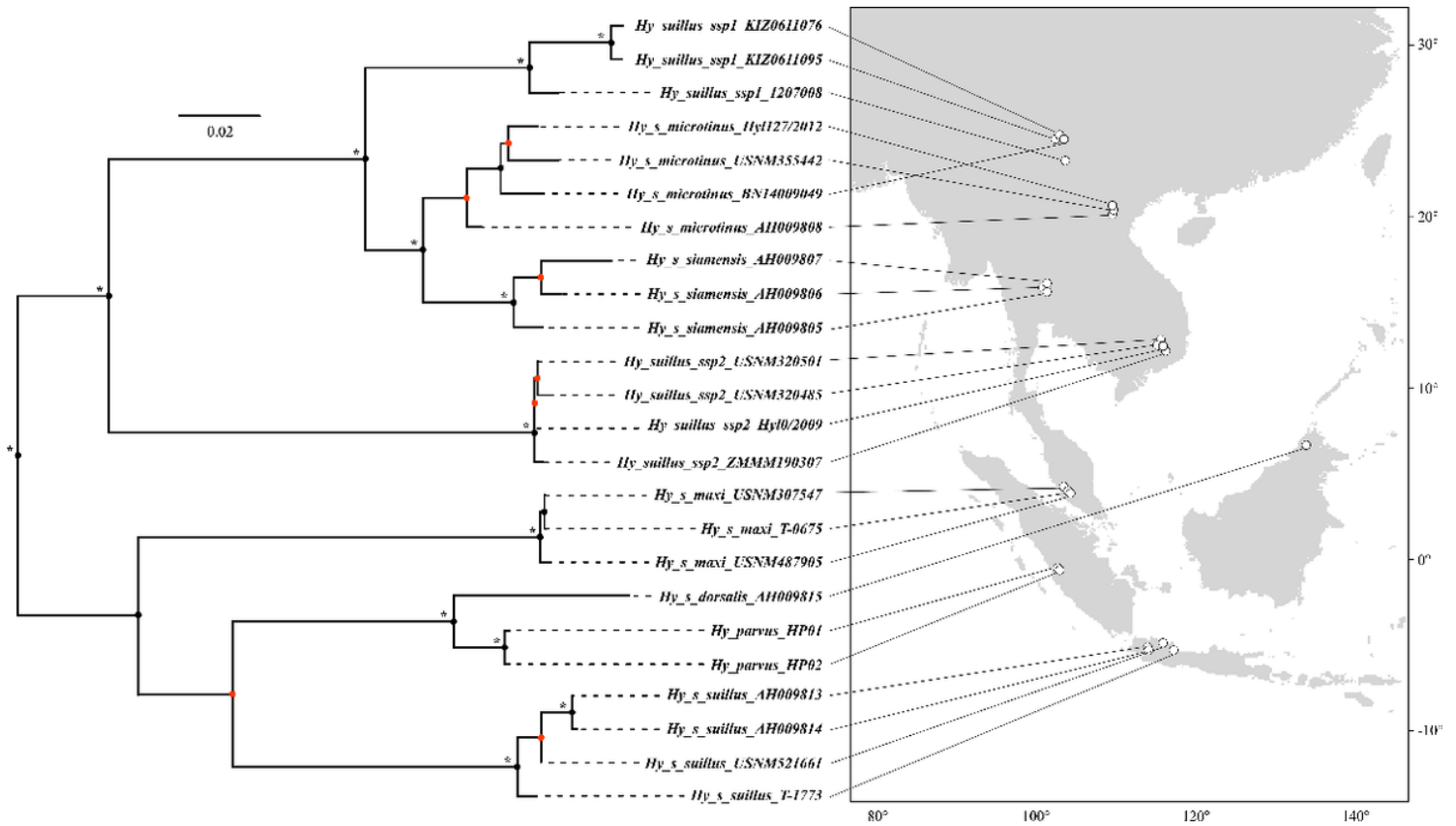


Figure 2

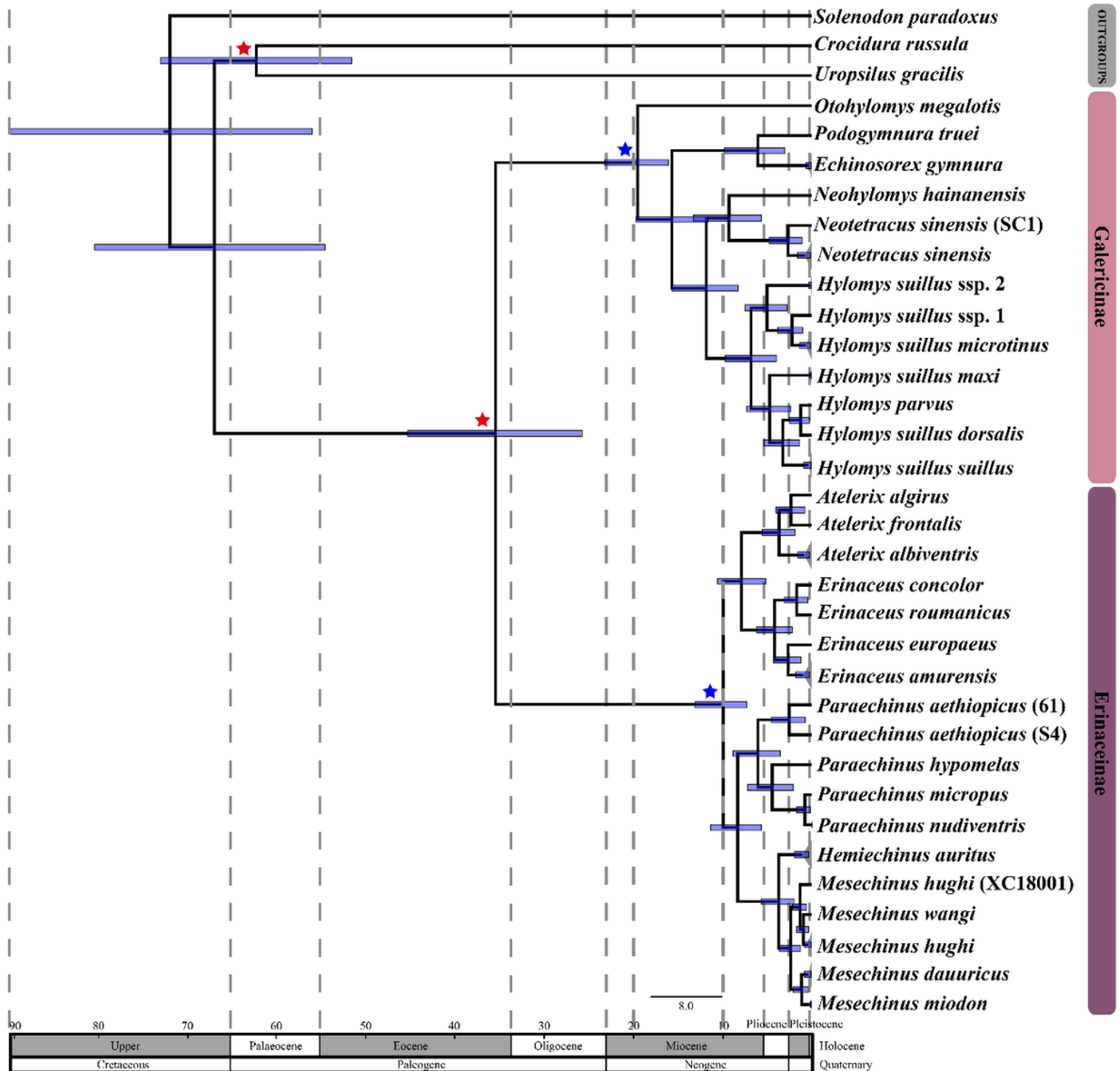
Bayesian tree of Erinaceidae based on MtG dataset. A red node indicates that Bayesian post probability (PP) of the node is lower than 70 (\* for ≥ 95). A node inside the red box was constrained. Multiple samplings of the same species or subspecies were collapsed. The name in the parentheses is an individual of the clade, which didn't cluster with other individuals of the same species for some reason.

The species that only possess the morphological signals was displayed as grey name and a name labelled † is fossil species.



**Figure 3**

Maximum likelihood tree of gymnures' genus *Hylomys* sampled in the analysis and their corresponding distributions. A node with an asterisk indicates the bootstrap proportion of the node is higher than or equal to 95. The remaining nodes that didn't label indicate the bootstrap proportion is between 70 and 95. The red node indicates the bootstrap proportion of the node is lower than 70.



**Figure 4**

Chronogram from the Bayesian analysis of MtG data using a relaxed molecular clock. Branch lengths represent time. Node bars indicate the 95% CI for the clade age. A node with a red pentagon indicates using a fossil calibration there and a node with a blue pentagon indicates using a secondary calibration there. The name in the parentheses is an individual of the clade, which did not cluster with other individuals of the same species for some reason.

## Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [3SupplementaryfiguresErinaceidae.docx](#)
- [4SupplementarytablesErinaceidae.xlsx](#)