Myoglobin primary structure reveals multiple convergent transitions to semi-aquatic life in the world's smallest mammalian divers

Kai He,[*,1,2,3] Triston G. Eastman,[1] Hannah Czolacz,[4] Shuhao Li,[1] Akio Shinohara,[5] Shin-ichiro Kawada,[6] Mark S. Springer,[7] Michael Berenbrink,[*,†,4] and Kevin L. Campbell[*,†,1]

[1]Department of Biological Sciences, University of Manitoba, Winnipeg, MB, Canada
[2]School of Basic Medical Sciences, Southern Medical University, Guangzhou, Guangdong, China
[3]State Key Laboratory of Genetic Resources and Evolution, Kunming Institute of Zoology, Chinese Academy of Sciences, Kunming, Yunnan, China
[4]Department of Evolution, Ecology and Behaviour, University of Liverpool, Liverpool, United Kingdom
[5]Department of Bio-resources, Division of Biotechnology, Frontier Science Research Center, University of Miyazaki, Miyazaki, Japan
[6]Department of Zoology, Division of Vertebrates, National Museum of Nature and Science, Tokyo, Japan
[7]Department of Evolution, Ecology, and Organismal Biology, University of California, Riverside, CA, USA

*Corresponding authors: Email: hekai2018@smu.edu.cn, Michael.Berenbrink@liverpool.ac.uk, or Kevin.Campbell@umanitoba.ca
[†]These authors contributed equally to this work

Abstract

Identifying the phylogenomic underpinnings of specialized phenotypes that fueled transitions into new adaptive zones is central to evolutionary biology but is often confounded by a fragmentary fossil record, morphological convergence, and unresolved phylogenetic relationships. The speciose mammalian order Eulipotyphla (e.g., moles, shrews, hedgehogs, solenodons) combines an unusual diversity of semi-aquatic, semi-fossorial, and fossorial forms that arose from terrestrial forbearers, yet the ecomorphological pathways leading to these lifestyles have been disputed for a century and more, calling for novel approaches. Here we resolve previously intractable eulipotyphlan intra-family relationships and establish the net surface charge of the oxygen-storing muscle protein myoglobin-readily determined from its primary structure-as a molecular signature to trace ancient lifestyle transitions based on protein sequence alone. Our analyses confidently resolve fossorial habits having evolved twice in talpid moles and reveal five independent origins of a semi-aquatic lifestyle in the order housing the world's smallest endothermic divers.

Introduction

A fundamental challenge of evolutionary biology is to understand the phylogenomic foundations of biochemical and physiological specializations that have enabled organisms to proliferate into new adaptive zones. While species of most mammalian orders generally occupy a single niche, members of a few mammalian lineages have radiated into a diverse range of environmental habitats. The order Eulipotyphla is one such assemblage, in that terrestrial forms repeatedly evolved into high-altitude, scansorial (climbing), fossorial (subterranean), semi-fossorial, and semi-aquatic niches (Burgin and He 2018; He, et al. 2017).

The fourth largest among the 27 mammalian orders after Rodentia, Chiroptera, and Cetartiodactyla, the eulipotyphlan clade consists of 527 recognized species of small insectivorous mammals distributed among four extant families (Burgin, et al. 2018): Erinaceidae (hedgehogs, gymnures, and moonrats), Soricidae (shrews), Talpidae (moles, shrew moles, shrew-like moles, and desmans), and Solenodontidae (solenodons). A fifth family (Nesophontidae, Caribbean 'island-shrews') only became extinct in the past 500 years following the arrival of Europeans in the West Indies (MacPhee, et al. 1999). Fossil evidence supports terrestrial habits as ancestral for Eulipotyphla, with shrew-like moles, erinaceids, and the majority of shrews

retaining morphological characteristics (e.g., slender legs and a long, pointed mouse-like snout; fig. 1) effective for hunting and avoiding predation on land (Churchfield 1990; Nowak 1999). Semi-aquatic moles and shrews possess a stream-lined body and limb adaptations for underwater locomotion (Nowak 1999; Burgin and He 2018), with desmans having an outward appearance more similar to that of semi-aquatic rodents (e.g., muskrat) than to that of any other eulipotyphlan mammal (fig. 1). By contrast, fossorial moles have evolved powerful forelimbs that are held out to the sides of their tube-shaped bodies and have large, spade-like hands facing posteriorly and oriented vertically allowing them to be effective burrowers (Gorman and Stone 1990). Finally, semi-fossorial shrews, solenodons, and shrew moles exhibit specializations in-between that of terrestrial and fossorial lineages, often retaining pointed mouse-like snouts yet possessing forelimbs that are oriented downwards and backwards to facilitate digging (Gorman and Stone 1990; Nowak 1999).

Fig. 1. Museum specimen photos illustrating the four major ecomorphotypes within the order Eulipotyphla. Representative terrestrial (shrew-like mole, *Uropsilus soricipes*; bottom right), semi-aquatic (Russian desman, *Desmana moschata*; right centre), strictly fossorial (Eastern mole, *Scalopus aquaticus*; left), and semi-fossorial (Chinese long-tailed mole, *Scaptonyx fusicaudus*; top centre) talpid mole species are given. Photo by Kai He.

While the lifestyle of extant eulipotyphlan mammals is easily predicted based on external morphological features alone (e.g., Woodman and Gaffney 2014; Sansalone, et al. 2016; Sansalone, et al. 2019), definitive assignments of transitional evolutionary stages have remained elusive due to an incomplete, fragmentary fossil record (Hickman 1984; Sánchez-Villagra, et al. 2006; Piras, et al. 2012; Hooker 2016). Additionally, because morphological specializations coincide with lifestyle across the different clades of eulipotyphlan mammals, it has been difficult to discern whether fossorial and semi-aquatic specializations found in Eurasian and North American moles arose in a common ancestor or are due to convergent evolution (Schwermann and Thompson 2015; He, et al. 2017). As a result, long-standing questions persist regarding the evolutionary pathways leading to the diverse ecomorphological variation seen in eulipotyphlan mammals.

For example, examinations of shoulder anatomy (Campbell 1939) and skeletal musculature (Whidden 1999) indicated that the shoulder and humeroclavicular joint morphology of fossorial talpid species aligned more closely with semi-aquatic locomotion, suggesting that the ancestor of Talpinae moles was semi-aquatic, which we here term the 'aquatic mole' hypothesis. Conversely, Reed (1951) analyzed anatomical features from four talpid genera and proposed that semi-fossorial and fossorial forms evolved from a terrestrial ancestor and that the semi-aquatic lifestyle of desmans was secondary to a semi-fossorial phase, a view also shared by Grand et al. (1998). Skeletal and fossil evidence also suggested that a period of strictly fossorial life preceded the invasion of semi-aquatic habits by the ancestors of star-nosed moles (*Condylura cristata*) (Hickman 1984; Grand, et al. 1998; Sanchez-Villagra, et al. 2006; Piras, et al. 2012; Hooker 2016), though Sansalone et al. (2016) proposed that the semi-aquatic lifestyle of this genus instead represents an autapomorphy from a semi-fossorial (shrew mole-like) ancestor.

Adaptive morphology and behavioural/physiological specializations for diving have also been well documented for members of the family Soricidae (Hutterer 1985; Catania, et al. 2008; Burgin and He 2018), though uncertainties regarding the evolution of semi-aquatic habits in this clade are also pervasive. For example, semi-aquatic members of four genera, *Sorex* (North American water shrews), *Neomys* (European water shrews), *Chimarrogale* (Asian water shrews), and *Nectogale* (Elegant water shrews) show a graded series of progressively greater morphological adaptations for underwater foraging (Hutterer 1985), yet phylogenetic analyses strongly refute a common ancestry of aquatic exploitation (He et al. 2010). Ancestral-state

83 reconstructions and the placement of *Chimarrogale* + *Nectogale* as sister-taxa instead provided
84 support for semi-aquatic habits having evolved three times (He et al. 2010), as opposed to four
85 convergent invasions (Churchfield 1990). However, based on associated fauna/flora and
86 paleoecological indicators of one of the oldest fossil nectogaline genera (*Asoriculus*), it has been
87 hypothesized that a shared Late Miocene ancestor of *Neomys*, *Chimarrogale*, and *Nectogale* may
88 have been adapted to humid environments with permanent open water (Rofes and Cuenca-
89 Bescos 2006), thus allowing the possibility that semi-aquatic habits evolved only twice in
90 shrews. Taken together, potential morphological convergence and an incomplete fossil record
91 have precluded reliable conclusions to be made regarding the evolutionary pathways underlying
92 the origins of the diverse phenotypes of eulipotyphlan mammals.
93 The monomeric oxygen-binding protein myoglobin, which plays an essential role in O_2
94 storage and facilitates O_2 diffusion within the heart and skeletal muscle of most vertebrates,
95 provides an intriguing, objective molecular target to address this question. Indeed, the O_2 storage
96 function of myoglobin has long been known to contribute integrally to the mammalian dive
97 response, with maximum active submergence times of birds and mammals being strongly
98 correlated to muscle myoglobin concentrations (Noren and Williams 2000; Lestyk, et al. 2009;
99 Ponganis, et al. 2011). Links between the molecular evolution of this molecule and aquatic life
100 had already been suspected for a half century or more (Scholander 1962; Romero-Herrera, et al.
101 1973), leading to the speculation that "there might be functional reasons, perhaps associated with
102 diving" underlying the presence of parallel amino acid replacements in the myoglobins of seals
103 and cetaceans (Romero-Herrera, et al. 1978). Mirceta et al. (2013) extended this work to reveal
104 that maximal muscle myoglobin concentration was mechanistically linked to myoglobin net
105 surface charge (Z_{Mb}) in mammals via adaptive changes in primary structure, with convergent
106 increases in Z_{Mb} found in members of all eight lineages with an extended aquatic/semi-aquatic
107 evolutionary history. This trait presumably represents an adaptive response to combat the general
108 propensity for proteins to precipitate at high concentration, thereby allowing for advantageous
109 elevations of this muscle O_2 store without deleterious self-aggregation (Mirceta, et al. 2013). An
110 increase in Z_{Mb} may also reduce the tendency of the myoglobin peptide (apomyoglobin) to
111 partially unfold prior to heme insertion, thus limiting hydrophobic interactions and making the
112 formation of an aggregate to be energetically unfavourable (Samuel, et al. 2015; Isogai, et al.
113 2018). It should be stressed that increases in myoglobin concentration afforded by an elevated

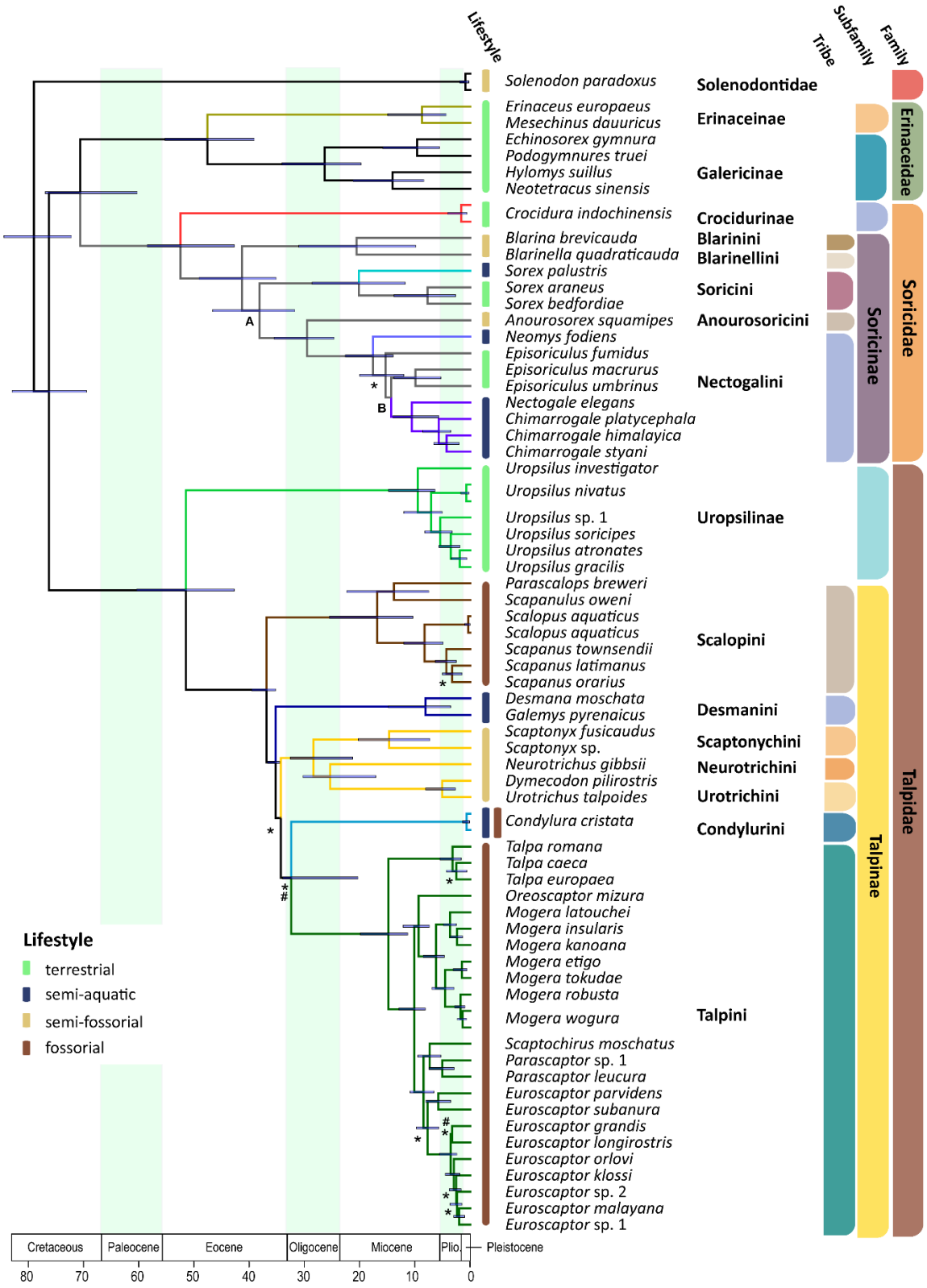
114 Z_{Mb} presumably represent an evolutionary trade off because there is less room for contractile
115 units (sarcomeres) and organelles (e.g., mitochondria) in muscle tissue as more energy and space
116 is allocated to myoglobin. Therefore, an increased Z_{Mb} only offers an advantage to species
117 experiencing hypoxic episodes as for example occurs during breath-hold dives when access to
118 environmental air is impeded. Importantly, the finding that Z_{Mb} and maximal myoglobin
119 concentrations are only slightly elevated within mammals that live at high elevations or have
120 fossorial lifestyles (Mirceta, et al. 2013) suggests that this locus may be useful to discern past
121 semi-aquatic versus high-altitude/fossorial evolutionary histories within eulipotyphlan mammals.
122 Support for the above contention is provided by the molecular modelling and phylogenetic
123 reconstruction of Z_{Mb} in six eulipotyphlan species (Mirceta, et al. 2013), which revealed
124 convergent elevations in Z_{Mb} in both semi-aquatic taxa included in the dataset—the American
125 water shrew (*Sorex palustris*) and the star-nosed mole. While myoglobin concentration data is
126 sparse for eulipotyphlan mammals, available data indicates that myoglobin levels are also higher
127 in these two species relative to non-aquatic forms (McIntyre, et al. 2002; Gusztak 2008).
128 Notably, Z_{Mb} was reconstructed to secondarily decline in the branch containing the single strictly
129 fossorial talpid species (Mirceta, et al. 2013), thus providing tentative support for the 'aquatic
130 mole' hypothesis. The sparse taxon sampling in the cited study (two moles, three shrews, and a
131 hedgehog), however, did not account for the broad phylogenetic and ecomorphological diversity
132 within these families. Additionally, current phylogenetic hypotheses for Eulipotyphla lack
133 definitive resolution below the family level (He, et al. 2010; He, et al. 2017) thereby precluding
134 reliable ancestral reconstructions. To overcome these shortcomings, we used a capture
135 hybridization approach to target coding sequences of myoglobin together with 25 tree-of-life
136 genes from 61 eulipotyphlan DNA libraries (44 moles, 11 shrews, 5 hedgehogs, and 1
137 solenodon) that included representatives from all seven recognized semi-aquatic genera within
138 this order. We then tested whether Z_{Mb} is elevated within members of all living genera of semi-
139 aquatic Eulipotyphla, and if so, whether there is a significant correlation between this lifestyle
140 and Z_{Mb}. Having shown that these important conditions are met, we then traced Z_{Mb} across
141 eulipotyphlan evolutionary history, thereby allowing us to determine when and how many times
142 semi-aquatic specializations for increased dive durations evolved in both shrews and moles, and
143 evaluate the merit of the 'aquatic mole' hypothesis versus alternative evolutionary scenarios of
144 talpid lifestyle evolution.

Results

Phylogenetic relationships within Eulipotyphla

To obtain mammalian tree-of-life genes (Meredith, et al. 2011) for phylogenetic estimation, we conducted in-solution probe-hybridization for segments of 25 single-copy genes from 61 eulipotyphlan DNA libraries (supplementary table S1). Twenty-three of these loci were efficiently captured, with usable sequence obtained from 51 to 61 libraries per locus, equivalent to a ~95% (1330/1403) success rate (supplementary table S2). By contrast, probes for two gene segments only successfully hybridized to 21 and 15 of the 61 libraries, respectively, and were subsequently not included in the phylogenetic analyses. Specifically, we failed to capture the prepronociceptin (PNOC) gene for any shrew and many of the talpid species (supplementary table S2). Probes for the interphotoreceptor retinoid binding protein (IRBP) also failed to retrieve any sequence from the solenodon and 37 of the 38 non-uropsiline talpid specimens. Notably, the putative 234 bp IRBP fragment recovered from True's shrew mole (*Dymecodon pilirostris*) contained gaps and premature stop codons, suggesting this gene was inactivated in the common ancestor of the subfamily Talpinae. Similarly, the Chinese mole shrew (*Anourosorex squamipes*) IRBP sequence contained premature stop codons and was presumed to be non-functional. After incorporating orthologous tree-of-life sequence data from ten additional eulipotyphlan specimens and five outgroup species downloaded from GenBank (see Methods for details), our 76 specimen dataset (71 eulipotyphlans) resulted in a final alignment of 39,414 bp.

Fig. 2. Time calibrated Bayesian phylogenetic tree of Eulipotyphla based on a concatenated alignment of 23 nuclear genes (outgroups not shown). Branch lengths represent median ages. Node bars indicate the 95% confidence interval [CI] for each clade age. Unless specified, all relationships are highly supported. Relationships weakly supported in concatenation Bayesian and maximum likelihood (PP <0.97 and/or BS <80: #) as well as *BEAST and ASTRAL coalescent analyses (C-PP <0.97 and/or C-BS <80: *) are indicated. An alternative position was recovered for Soricini in the ASTRAL tree (supplementary fig. S2), while both coalescent analyses (ASTRAL, *BEAST) favored *Episoriculus* monophyly (supplementary figs. S2, S3). (next page)

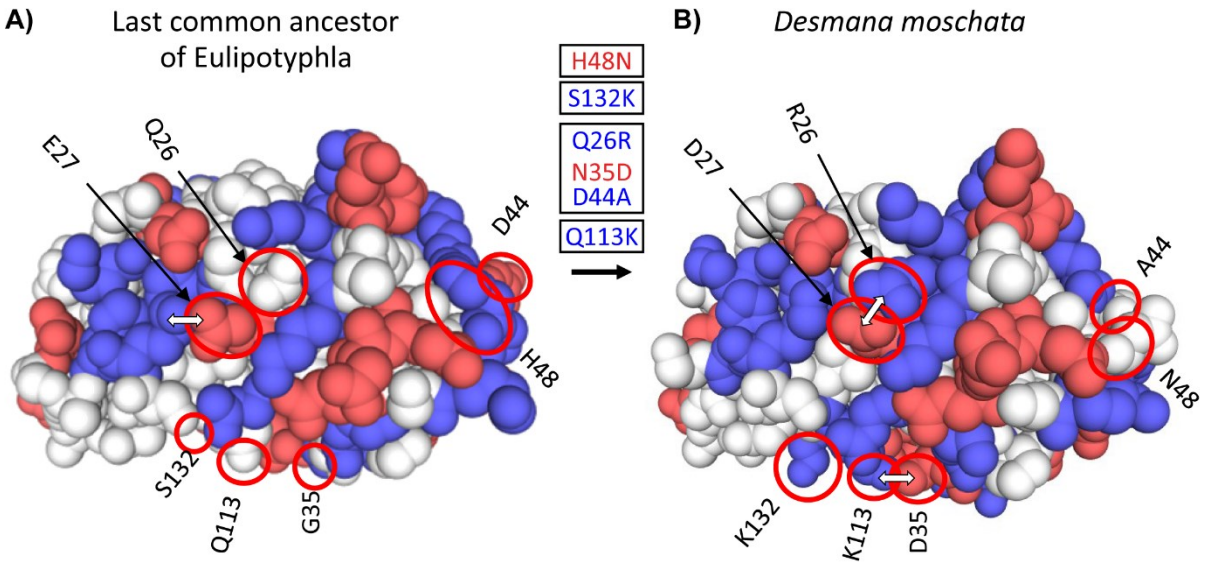

Lifestyle
Tribe
Subfamily
Family
Solenodon paradoxus
Solenodontidae
Erinaceus europaeus
Mesechinus dauuricus
Echinosorex gymnura
Podogymnures truei
Hylomys suillus
Neotetracus sinensis
Erinaceinae
Galericinae
Erinaceidae
Crocidura indochinensis
Crocidurinae
Blarina brevicauda
Blarinella quadraticauda
Blarinini
Blarinellini
Sorex palustris
Sorex araneus
Sorex bedfordiae
Soricini
Anourosorex squamipes
Anourosoricini
Neomys fodiens
Episoriculus fumidus
Episoriculus macrurus
Episoriculus umbrinus
Nectogalini
Nectogale elegans
Chimarrogale platycephala
Chimarrogale himalayica
Chimarrogale styani
Uropsilus investigator
Uropsilus nivatus
Uropsilus sp. 1
Uropsilus soricipes
Uropsilus atronates
Uropsilus gracilis
Uropsilinae
Soricinae
Soricidae
Parascalops breweri
Scapanulus oweni
Scalopus aquaticus
Scalopus aquaticus
Scapanus townsendii
Scapanus latimanus
Scapanus orarius
Scalopini
Desmana moschata
Galemys pyrenaicus
Desmanini
Scaptonyx fusicaudus
Scaptonyx sp.
Scaptonychini
Neurotrichus gibbsii
Dymecodon pilirostris
Neurotrichini
Urotrichus talpoides
Urotrichini
Condylura cristata
Condylurini
Talpa romana
Talpa caeca
Talpa europaea
Oreoscaptor mizura
Mogera latouchei
Mogera insularis
Mogera kanoana
Mogera etigo
Mogera tokudae
Mogera robusta
Mogera wogura
Talpini
Scaptochirus moschatus
Parascaptor sp. 1
Parascaptor leucura
Euroscaptor parvidens
Euroscaptor subanura
Euroscaptor grandis
Euroscaptor longirostris
Euroscaptor orlovi
Euroscaptor klossi
Euroscaptor sp. 2
Euroscaptor malayana
Euroscaptor sp. 1
Talpinae
Talpidae
A
B
Lifestyle
terrestrial
semi-aquatic
semi-fossorial
fossorial
Cretaceous
Paleocene
Eocene
Oligocene
Miocene
Plio.
Pleistocene
80
70
60
50
40
30
20
10
0

174

175 To estimate evolutionary relationships, we used maximum likelihood (RAxML) and
176 Bayesian (BEAST) approaches on the concatenated alignment, and coalescent-based species tree
177 methods (ASTRAL-III, *BEAST) on the 23 individual gene trees. These analyses resulted in
178 highly congruent topologies (fig. 2 and supplementary figs. S1, S2, S3), with family level
179 relationships corresponding to those obtained in recent eulipotyphlan (Brace, et al. 2016;
180 Springer, et al. 2018) and mammal wide studies (Meredith, et al. 2011; Esselstyn, et al. 2017).
181 Higher-level relationships within Soricidae and Erinaceidae also corresponded closely with
182 previous studies on these families (He, et al. 2010; He, et al. 2012). Principal among these is the
183 non-sister relationship of the Nectogalini water shrew clades *Neomys* and *Chimarrogale* +
184 *Nectogale*, as the latter was strongly supported as sister to terrestrial *Episoriculus* by RAxML
185 (bootstrap support [ML-BS] = 97; supplementary fig. S1), BEAST (posterior probability [PP] =
186 1.0; fig. 2), ASTRAL-III (coalescent bootstrap support based on gene-wise resampling [C-BS] =
187 76; supplementary fig. S2), and *BEAST (posterior probability [C-PP] = 0.99; supplementary
188 fig. S3). By contrast, previously intractable relationships within Talpidae are now resolved, with
189 the interrelationships among fossorial and semi-aquatic clades consistently recovered (fig. 2 and
190 supplementary figs. S1, S2, S3). Specifically, desmans are placed sister to a clade containing
191 shrew moles, star-nosed moles, and Eurasian fossorial moles (BS = 100, PP = 1.0, C-BS = 87, C-
192 PP = 0.97), with Condylurini and Talpini recovered as sister lineages (BS = 72, PP = 1.0, C-BS
193 =52, C-PP = 0.93). North American fossorial Scalopini moles are also supported as sister to all
194 other Talpinae moles with high support scores (BS = 97, PP = 1.0, C-BS = 87, C-PP = 0.97).
195 Accordingly, a sister group relationship of fully fossorial Scalopini and Talpini moles was
196 statistically rejected by the Shimodaira–Hasegawa (SH) test ($p < 0.01$, supplementary table S3).
197 The only minor incongruences among phylogenies was the position of the tribe Soricini and of
198 the Taiwanese brown-toothed shrew (*Episoriculus fumidus*) within Nectogalini in the ASTRAL
199 tree (fig. 2 and supplementary figs. S1, S2, S3).
200
201 *Myoglobin primary structure*
202 Complete myoglobin coding sequences (465 base pairs including initiation and stop codons)
203 were obtained for 55 eulipotyphlan species (38 moles, 14 shrews, 2 erinacids, and 1 solenodon)
204 using capture-hybridization, transcriptome sequencing, genome mining, and PCR approaches.
205 Conceptual translations revealed the expected 153 amino-acid peptides in most cases. However,

206 the translated myoglobin proteins of the Chinese mole shrew, both members of the genus
207 *Scaptonyx*, and the two desman genera are only 152 amino acids in length. In every case the
208 shorter myoglobin sequences are the result of 3 bp deletions in exon 3 that corresponded to
209 residue position 121 (supplementary table S4) and shorten the loop between helices G and H
210 from 6 to 5 residues. These deletions were confirmed for each lineage using at least two of the
211 three sequencing approaches noted above. To our knowledge, a similar deletion was previously
212 only known for three bird of prey species (Enoki et al. 2008), but we have detected it also in the
213 predicted myoglobin sequences of the draft genomes of several burrowing species of the order
214 Rodentia (data not shown), notably including the fully fossorial Transcaucasian and Northern
215 mole voles (*Ellobius lutescens* and *E. talpinus*, respectively; Mulugeta, et al. 2016).
216
217 *Myoglobin homology modelling*
218 Homology modelling of myoglobin structure was conducted for one extant species from
219 each of the five diving lineages together with that predicted for the last common eulipotyphlan
220 ancestor based on ancestral sequence reconstruction (see next section). This analysis confirmed
221 that all charge-changing substitutions were located at the solvent-exposed surface of the protein
222 (supplementary table S4). These comparisons further suggested that deletion of position 121 in
223 the loop between helices G and H in the myoglobin of the Russian desman has minimal effect on
224 the tertiary structure of the protein (supplementary fig. S5). In addition to their effect on Z_{Mb},
225 some of the charge-increasing substitutions were associated with a complex re-arrangement of
226 the network of salt bridges in the tertiary structure of the protein. Thus, for example, during the
227 evolution of the Russian desman, arguably the most aquatic of extant Eulipotyphla, the two
228 neighbouring substitutions Gln^{26}→Arg^{26} (charge-increasing) and Glu^{27}→Asp^{27} (charge-neutral)
229 in the B-helix allow for the formation of a new intra-helical salt bridge between Arg^{26} and Asp^{27},
230 while at the same time breaking the inter-helical salt bridge between Glu^{27} and Lys^{118} in the B-
231 and G-helices, respectively (fig. 3). However, charge-decreasing Asn^{35}→Asp^{35} followed by
232 charge-increasing Gln^{113}→Lys^{113} subsequently again tethered the B- and G-helices to each other
233 by the salt bridge Asp^{35}- Lys^{113}. Further, the removal of a negative charge by the substitution
234 Asp^{44}→Ala^{44} destroyed a salt bridge between Asp^{44} and Lys^{47} in the CD-corner of the protein,
235 thereby potentially affecting the flexibility of the loop between the C- and D-helices (fig. 3).
236 However, a detailed assessment of any changes in the folding stability of the proteins that are

237 associated with the identified charge-changing substitutions, let alone with any charge neutral
238 replacements (see, e.g., Isogai, et al. 2018), is difficult and beyond the scope of this study, not
239 least because of the potentially opposing effects of salt bridge formation and the associated
240 desolvation of charges for the folding stability of proteins (for discussion see Bosshard, et al.
241 2004).
242



243
244 **Fig. 3.** Three-dimensional structural models of myoglobin in (A) the last common ancestor of
245 Eulipotyphla and (B) the semi-aquatic Russian desman (*Desmana moschata*) obtained by
246 homology modelling using the SWISS-MODEL server (Waterhouse, et al. 2018). The structure
247 of the last common ancestor of the group was modelled based on results of an amino acid
248 sequence reconstruction (see text for details). Ancestral (left) and derived (right) states of charge-
249 changing amino acid replacements are circled and indicated with positional number and one-
250 letter amino acid code. Blue and red color indicate amino acids with positively (H, His; K, Lys;
251 R, Arg) and negatively charged amino acid side chains (D, Asp; E, Glu), respectively. White
252 double arrows indicate surface amino acid side chains involved in salt bridges that are affected
253 by charge-changing substitutions. Text boxes indicate the reconstructed temporal order (top to
254 bottom) of charge decreasing and charge increasing amino acid substitutions (red and blue font,
255 respectively) in the *Desmana* lineage in one letter code from ancestral (left) to derived (right)
256 separated by positional number. Note that charge neutral substitutions (e.g., G35N), are not given
257 in the text boxes.

Electrophoretic mobility and ancestral reconstruction of Z_{Mb}

Without exception, the modelled Z_{Mb} values of extant semi-aquatic taxa (2.07 to 3.07) were substantially higher than those of terrestrial Eulipotyphla (-0.46 to 0.63), with fossorial species generally exhibiting intermediate Z_{Mb} values (typically 1.07; fig. 4A and supplementary fig. S6). To assess the reliability of our Z_{Mb} determinations, we measured the electrophoretic mobility of the myoglobin band of muscle extracts from two semi-aquatic, two strictly fossorial, and one terrestrial eulipotyphlan species (fig. 4B). The close correspondence between the two variables validates Z_{Mb} as a molecular marker for inferring present and past semi-aquatic habits in Eulipotyphla.

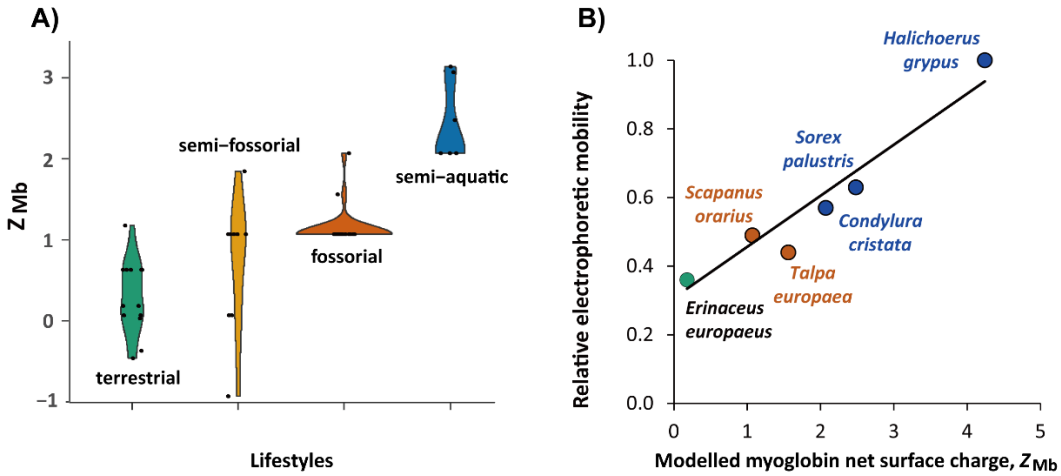


Fig. 4. (A) Violin plot showing the distribution (y-axis) and probability density (x-axis) of modelled myoglobin net surface charge, Z_{Mb}, among living species (black dots) of the four prevalent eulipotyphlan ecomorphotypes. (B) Correlation between Z_{Mb} and electrophoretic mobility of native myoglobin from five eulipotyphlan insectivores; data from the grey seal (*Halichoerus grypus*) is added for comparison. Z_{Mb} was calculated as the sum of the charge of all ionisable groups at pH 6.5 by modelling myoglobin primary structures onto the tertiary structure and using published, conserved, site-specific ionization constants (McLellan 1984; Mirceta, et al. 2013). Electrophoretic mobility was assessed relative to the mobility of grey seal myoglobin using native polyacrylamide gel electrophoresis of heart or skeletal muscle protein extracts of the indicated species. Green, orange, brown, and blue areas (A) or symbols and fonts (B) indicate terrestrial, semi-fossorial, fossorial, and semi-aquatic/aquatic species, respectively. Phylogenetic Generalized Least Squares analysis in panel (B) revealed a highly significant positive correlation ($R^2 = 0.897$, $p < 0.005$) between the two parameters (solid line, $y = 0.1488x + 0.3075$).

281

282 Our myoglobin nucleotide and amino acid gene trees retrieved few well-supported
283 phylogenetic relationships (especially at deeper nodes), and no compelling evidence of an
284 evolutionary history that deviated from the species trees (supplementary figs. S8A, B), which
285 would potentially result in erroneous ancestral reconstructions (Hahn and Nakhleh, 2016). We
286 thus conducted a maximum likelihood ancestral amino acid sequence reconstruction using the
287 species tree in fig. 1 as the phylogenetic backbone and the best fitting model of the Dayhoff
288 amino acid substitution matrix with a gamma distribution of rate variation among sites. This
289 analysis yielded highly supported ancestral amino acid identities across all 153 residues and 58
290 internal nodes of the species tree (supplementary fig. S7A), which was presumably due to our
291 dense taxon sampling and the relatively highly conserved primary structure of mammalian
292 myoglobins. Of the 8,874 (= 58 × 153) reconstructed ancestral sites on the species tree, 8,799
293 (99.15%) had maximal probabilities of reconstructed amino acid identities of $p > 0.95$ under the
294 given phylogeny and amino acid substitution model. In 50 cases (0.55%), alternatively
295 reconstructed amino acids with $p > 0.05$ were of the same charge and thus did not affect the
296 calculated Z_{Mb} values. Only in 25 cases (0.28%), one or more alternatively reconstructed amino
297 acids with $p > 0.05$ carried a different charge from the most probable amino acid at that site.
298 However, even in those cases this usually had only a minimal effect on Z_{Mb} (± 0.11; 3 sites), or
299 the summed probabilities of alternative charge states were comparatively small ($p \leq 0.21$; 23
300 sites) such that we regard both the results of the ancestral sequence reconstruction and of the
301 resulting values of Z_{Mb} as robust. This is supported by generally congruent results obtained by
302 codon-based ancestral sequence reconstructions (supplementary fig. S7B), which primarily
303 deviated from the above empirical amino acid matrix-based method at some deeper nodes, where
304 a higher number of negatively charged amino acid residues (and thus anomalously low values for
305 Z_{Mb}; e.g., -1.93) were reconstructed relative to those obtained for extant mammalian species
306 (Mirceta, et al. 2013; this study). Consequently, the following discussion largely focuses on the
307 Z_{Mb} reconstruction using the empirical Dayhoff substitution matrix (supplementary fig. S7A),
308 which moreover—in contrast to the codon-based analysis (supplementary fig. S7B)—includes
309 the effects of purifying selection acting on replacements with dissimilar amino acid properties.

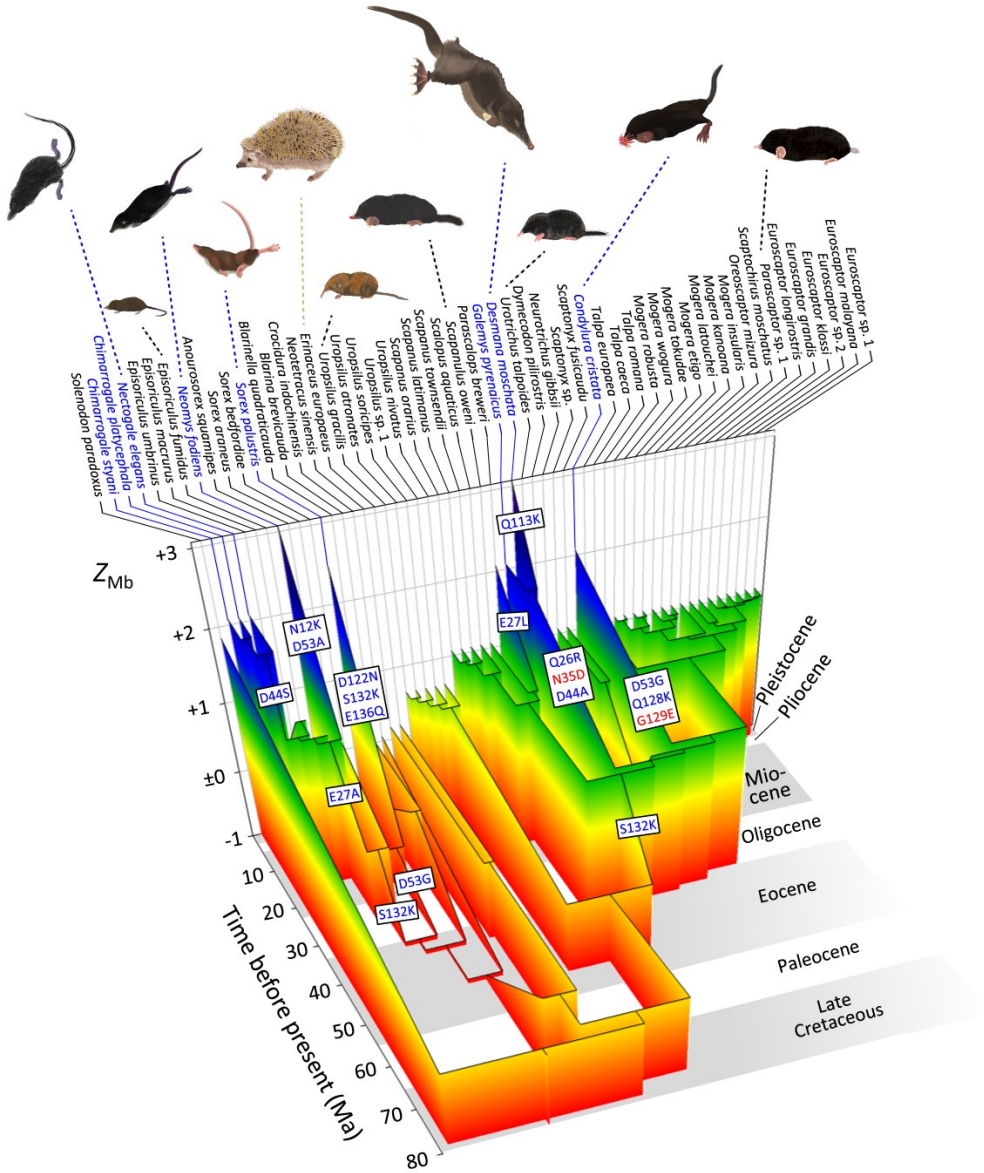


Fig. 5. Evolutionary reconstruction of myoglobin net surface charge Z_{Mb} in 55 eulipotyphlan insectivores mapped onto the time calibrated phylogeny of Fig. 2. Ancestral Z_{Mb} was modelled from primary structures as in Fig. 3 and after maximum likelihood ancestral sequence reconstruction. Major charge increasing (blue font) and charge decreasing (red font) amino acid substitutions, from ancestral to derived and separated by positional number, inferred for the immediate ancestry of semi-aquatic species (blue font) are indicated in textboxes alongside the respective branches. Grey and white background shading indicates geologic epochs. See supplementary Fig. S7A for a complete account of charge-changing substitutions, reconstructed Z_{Mb} values, and outgroup information. Paintings of representative species by Umi Matsushita.

320

321 Ancestral Z_{Mb} estimates arising from the best fitting amino-acid based reconstruction model
322 indicated that the most recent common ancestor of Eulipotyphla displayed a Z_{Mb} of 0.18. Within
323 Soricidae, distinct increases in Z_{Mb} were found on the branches leading to the three semi-aquatic
324 clades (fig. 5). Specifically, the Z_{Mb} of the American water shrew (*Sorex palustris*) branch
325 increased from -0.46 to 2.48 and was characterized by three integral (+1) charge increasing
326 residue replacements (Asp^{122}→Asn^{122}, Ser^{132}→Lys^{132}, and Glu^{136}→Gln^{136}; supplementary fig.
327 S4A). The other three genera of water shrews reside in the tribe Nectogalini, the stem branch of
328 which evolved a single charge increasing substitution (Glu^{27}→Ala^{27}) and thus had a Z_{Mb} of 1.07
329 (fig. 5). The European water shrew branch (*Neomys fodiens*) subsequently acquired two charge
330 increasing substitutions (Asn^{12}→Lys^{12}, Asp^{53}→Ala^{53}; Z_{Mb}=3.07; supplementary fig. S4B), while
331 the common ancestor of *Nectogale* + *Chimarrogale* evolved a separate charge increasing
332 replacement (Asp^{44}→Ser^{44}; Z_{Mb}=2.07; supplementary fig. S4C). By contrast, members of the
333 terrestrial genus *Episoriculus*, which are nested between the semi-aquatic nectogaline lineages,
334 exhibited secondary reductions in Z_{Mb} towards neutrality via different residue substitutions in
335 each case (supplementary fig. S7A). Importantly, similar results were obtained when the above
336 amino-acid based reconstructions were re-run on an alternative topology (supplementary fig. S3)
337 that supported the monophyly of *Episoriculus* (supplementary fig. S7C).
338

339 Within the Talpidae, Z_{Mb} exhibited an increase from an ancestral value of 0.07 to 1.07
340 (Ser^{132}→Lys^{132}) in the stem Talpinae branch (fig. 5), with the probabilities of the identities of
341 reconstructed amino acids at this node under the given tree and substitution model being 1.00 for
342 all 153 sites (supplementary fig. S7A). With few exceptions, these values remained highly
343 conserved in the clades containing either semi-fossorial or fossorial taxa (fig. 5 and
344 supplementary figs. S7A, B). By contrast, Z_{Mb} increased to >2 in members of both semi-aquatic
345 lineages—star-nosed moles and the desmans (fig. 3). For the star-nosed mole branch this entailed
346 two charge increasing substitutions (Asp^{53}→Gly^{53}, Gln^{128}→Lys^{128}) and one charge decreasing
347 substitution (Gly^{129}→Glu^{129}; supplementary figs. S4D and S7A). The ancestral branch of the
348 desmans also acquired two charge increasing replacements (Gln^{26}→Arg^{26}, Asp^{44}→Ala^{44})
349 together with one charge decreasing (Asn^{35}→Asp^{35}) replacement, with Z_{Mb} further being
350 elevated in the Russian desman branch via the acquisition of an additional charge increasing

351 substitution (Gln^{113}→Lys^{113}; Z_{Mb}=3.07; figs. 3, 5). Notably, while the same residue positions are
352 occasionally recruited in the charge altering replacements of semi-aquatic taxa, the derived
353 residues are different in all cases (supplementary table S5).
354 To evaluate the above reconstructions of semi-aquatic lifestyles based on Z_{Mb}, we coded each
355 species as semi-aquatic or non-aquatic and estimated ancestral lifestyles using both maximum
356 parsimony and threshold models. Although both of these latter analyses suggested that semi-
357 aquatic lifestyles evolved five times independently, this result was not strongly supported by the
358 threshold model (supplementary fig. S9A). For example, the posterior probabilities of the most
359 recent common ancestor of *Desmana* + *Galemys* and *Chimarrogale* + *Nectogale* being semi-
360 aquatic was only 0.85 and 0.75, respectively, while the posterior probability for the most recent
361 common ancestor of Nectogalini being semi-aquatic was 0.40. To account for the alternative
362 placement of *E. fumidus*, we repeated this analysis using the results of the *BEAST species tree
363 (supplementary fig. S3). The maximum parsimony reconstruction yielded two equi-parsimonious
364 ancestral reconstructions of semi-aquatic lifestyle in nectogaline shrews, encompassing a single
365 origin at the base of nectogaline shrews with a secondary loss at the base of the *Episoriculus*
366 clade (supplementary fig. S9B). By contrast, the threshold model only weakly supported two
367 independent origins of a semi-aquatic lifestyle in *Neomys* and *Nectogale* + *Chimarrogale*.
368
369 *Lifestyle correlation analyses*
370 As a final test regarding the reliability of using Z_{Mb} to predict ancient semi-aquatic lifestyles,
371 we first assigned species as semi-aquatic or non-aquatic (see fig. 2 and supplementary table S1
372 for lifestyle assignments), and used a threshBayes analysis to estimate covariances between Z_{Mb}
373 and a semi-aquatic lifestyle. This analysis revealed a strong correlation between Z_{Mb} and aquatic
374 adaptation (correlation coefficient r = 0.78, 95% highest posterior density [HPD] = 0.48-0.93;
375 supplementary fig. S10A). Conversely, threshBayes analyses did not support a correlation
376 between Z_{Mb} and adaptations for digging, or between Z_{Mb} and a fully fossorial habit, when
377 terrestrial and semi-aquatic eulipotyphlan species were included (supplementary figs. S10B, C).
378 When applying threshBayes to subsets of habits that included only terrestrial and fully fossorial
379 species, or terrestrial and burrowing species, weak correlations between Z_{Mb} and fossoriality (r =
380 0.55) and digging habits (r = 0.32) were revealed, but not significantly supported (i.e., 95%HPD
381 overlaps with 0; supplementary figs. S10D, E). Importantly, Z_{Mb} comparisons between semi-

aquatic and terrestrial, and between semi-aquatic and fossorial habits (supplementary figs. S10F, G) were both significant.

Discussion

The phylogenetic estimates constructed from our comprehensive tree-of-life gene set provide a robust framework to interpret ecomorphological evolution within Eulipotyphla. The close correspondence of our concatenation and coalescent phylogenetic topologies not only support key findings of previous studies—i.e., the monophyly of shrew moles, the non-monophyly of nectogaline water shrews (Whidden 2000; He, et al. 2010; He, et al. 2017)—but finally puts to rest the long hypothesized monophyletic origin of the fully fossorial tribes Talpini and Scalopini, which have been routinely grouped together based on morphological data (see, e.g., Whidden 2000; Piras, et al. 2012; Schwermann and Thompson 2015; Hooker 2016; Sansalone, et al. 2019). Results of the present study thus provide compelling evidence that the extreme anatomical specializations for subterranean life evolved convergently in these two tribes. The molecular phylogenetic position of the amphibious desmans and semi-aquatic/fossorial star-nosed mole are also finally resolved (fig. 2), with the latter placed sister to Talpini, a relationship not supported by morphological-based hypotheses (see, e.g., Whidden 2000; Motokawa 2004; Sanchez-Villagra, et al. 2006).

Previous studies have failed to reach consensus on the lifestyle evolution of Eulipotyphla. Here we show that ancestral sequence reconstruction of myoglobin primary structure and Z_{Mb} modelling, with their well established mechanistic and biophysical underpinnings, outperform discrete, two character-state lifestyle reconstructions based on maximum parsimony and threshold approaches. These attributes, together with demonstrated links between Z_{Mb} and maximal Mb concentration, and hence muscle oxygen storage capacity (Mirceta, et al. 2013; Berenbrink 2020), provide strong support that this quantitative metric is well suited to resolve long-standing questions regarding lifestyle evolution within Eulipotyphla. For example, despite being less specialized for aquatic life than marine mammals, the strong positive correlation between Z_{Mb} and semi-aquatic specialization indicates that Z_{Mb} is a powerful marker to identify secondary aquatic transitions in even the world's smallest mammalian divers. Although a strong correlation was not recovered between Z_{Mb} and strictly fossorial habitation, Z_{Mb} is highly conserved in fossorial Scalopini and Talpini, with 23 of 25 species exhibiting a value of 1.07.

413 This conservation is presumably driven by selective pressures to maintain moderately elevated
414 tissue myoglobin levels to help foster burst burrowing activities in their hypoxic underground
415 environment. By contrast, the Z_{Mb} of terrestrial species in our dataset were consistently close to
416 neutrality, though many lineages exhibited clear signals of Z_{Mb} fluctuation over time
417 (supplementary figs. S7A, B, C). For example, the shrew gymnure (*Neotetracus sinensis*) branch
418 evolved nine charge altering residue substitutions since its split from European hedgehogs
419 (*Erinaceus europaeus*), including the charge inversion Asp^{126}→Lys^{126} that increases Z_{Mb} by +2.
420 Similarly, the Taiwanese brown-toothed shrew branch (*Episoriculus fumidus*) has fourteen
421 charge altering residue substitutions, including one negative-to-positive (Glu^{109}→Lys^{109}) and two
422 positive-to-negative charge inversions (Lys^{102}→Glu^{102} and Lys^{132}→Glu^{132}). These observations
423 are consistent with a stochastic evolutionary process operating under purifying selection. Charge
424 fluctuation is also apparent on the Hispaniolan solenodon (*Solenodon paradoxus*) branch, with
425 the moderately elevated Z_{Mb} value (1.85) in line with their terrestrial/burrowing lifestyle (Nowak,
426 1999). Myoglobin sequence data from additional extant/extinct members of this family (e.g.,
427 Cuban solenodons) together with that from the recently extinct terrestrial/fossorial Antillean
428 family Nesophontidae (MacPhee, et al. 1999)—which is placed sister to solenodons (Brace et al.
429 2016)—may help resolve the life history evolution of this poorly understood insectivore clade.
430 Similar, albeit less pronounced charge fluctuations are evident in several fossorial mole branches
431 (as well as the stem *Condylura* and desman branches), suggesting that the single fossorial Z_{Mb}
432 outlier (hairy-tailed mole, *Parascalops breweri*; Z_{Mb} = 2.07) may also represent stochastic
433 variation that has not yet been selected against and that presumably will return to an ecologically
434 normalized value over evolutionary time.
435 Our results indicate that charge increases in Z_{Mb} to >2 is essential for members of
436 eulipotyphlan mammals to successfully exploit a semi-aquatic lifestyle. Because Z_{Mb} of the most
437 recent common ancestor of Talpinae was confidently estimated to be below this value (fig. 5),
438 our results do not support the "aquatic mole" hypothesis which posits that a semi-aquatic stage
439 predated the invasion of fossorial habits in talpid moles (Campbell 1939; Whidden 1999). This
440 conclusion is supported by the results of our lifestyle reconstructions, which uniformly revealed
441 that stem Talpinae had a low probability of semi-aquatic habits. These findings are instead
442 consistent with the interpretation that fossorial forms evolved directly from terrestrial/semi-
443 fossorial ancestors, without passing through a semi-aquatic phase (Reed 1951; Hickman 1984;

444 Grand, et al. 1998). The interpretation of a semi-fossorial ancestry for early non-uropsiline
445 (Talpinae) moles is further supported by our finding of a presumed inactivation/deletion of IRBP
446 on this branch (supplementary Table S2). This eye-specific locus has been shown to be
447 inactivated/lost in numerous mammalian lineages that inhabit subterranean and other dim-light
448 niches (Emerling and Springer 2014), and also appears to be non-functional in solendons and the
449 Chinese mole shrew (this study). The Z_{Mb} charge increase (to 2.07) in the stem desman branch,
450 paired with a lack of pronounced forelimb specializations for digging (He, et al. 2017) and a
451 basal placement among non-scalopine Talpinae (fig. 2), also chimes with Reed's (1951)
452 suggestion that the semi-aquatic lifestyle of this clade was secondary to a semi-fossorial phase.
453 By contrast, semi-aquatic/fossorial star-nosed mole exhibits prominent morphological
454 adaptations for burrowing and is placed sister to fossorial Talpini, consistent with Grand et al.'s
455 (1998) hypothesis that this lineage passed through a specialized fossorial stage prior to invasion
456 of the semi-aquatic niche. However, a semi-fossorial—as opposed to a fully fossorial—ancestry
457 for *Condylura* (Sansalone, et al. 2016) cannot be excluded based on the pattern of Z_{Mb} evolution
458 on this branch.

459 The charge elevation to 1.07 in stem Nectogalini shrews is enticing, as it temporally
460 corresponds to the early Miocene fossil *Asoriculus*, which was theorized to have inhabited wet
461 environments though was unlikely to have been an efficient semi-aquatic predator (Rofes and
462 Cuenca-Bescos 2006). The single charge increasing substitution that evolved at this stage
463 ($Asp^{27} \rightarrow Ala^{27}$) is retained in the three semi-aquatic nectogaline genera and presumably
464 facilitated the adaptation of early *Neomys* and *Chimarrogale* + *Nectogale* for aquatic food
465 resources starting some 15 and 10 million years ago, respectively. The presence of separate
466 charge increasing replacements within the evolutionary branches of each of the three
467 *Episoriculus* species also opens the possibility of additional semi-aquatic 'experiments' in this
468 genus. Regardless, independent reductions in Z_{Mb} to neutrality or below for each of the latter
469 extant species, which today inhabit damp areas in vegetated environments (Nowak 1999), is
470 consistent with convergent reversions of *Episoriculus* lineages to a predominantly terrestrial
471 foraging habit.

472 Our results provide strong evidence that the exploitation of semi-aquatic habits by extant
473 shrews and talpids occurred at least five times, and was accompanied by convergently evolved
474 charge-increasing substitutions at different surface sites on the myoglobin protein in each case

475 (figs. 3 and 5 and supplementary figs. S4, S7). This finding provides additional support for our
476 contention that adaptive increases in Z_{Mb} underlie the invasion of (semi-)aquatic niches by
477 mammals, presumably by allowing for higher skeletal muscle myoglobin concentration (Mirceta,
478 et al. 2013; Berenbrink 2020). This elevated Z_{Mb} presumably underlies the elevated O_2 reserves
479 in muscle of star-nosed moles and American water shrews compared to non-aquatic relatives
480 (McIntyre, et al. 2002; Gusztak 2008), and likely contributes to the extended dive times and
481 remarkable underwater foraging efficiency of these species (Catania, et al. 2008). The increase in
482 Z_{Mb} must be particularly important for semi-aquatic soricine shrews due to allometic
483 considerations that have resulted in extremely high muscle mitochondrial contents (which may
484 comprise up to 45% of the cell volume; Weibel 1985) and mass-specific tissue O_2 requirements
485 that may be >100-fold higher than those of large-bodied marine mammals (Butler 1998; Gusztak,
486 et al. 2005). It is notable that the highest Z_{Mb} values were found for Russian desmans (3.07) and
487 European water shrews (3.07), consistent with the exceptional diving abilities of these species
488 (Vogel 1998; Ivlev 2010), and it is predicted that these two species will also possess the highest
489 muscle myoglobin concentrations among Eulipotyphla.
490 Adaptive evolution of similar phenotypic and physiological features occurring in distantly
491 related lineages are not uncommon in mammals (Madsen, et al. 2001). For example, adaptive
492 radiations in Afrotheria and Laurasiatheria resulted in striking morphological convergence of
493 species occupying semi-aquatic (otter shrews vs. desmans) and subterranean (golden moles vs.
494 true moles) habitats (Madsen, et al. 2001; Springer, et al. 2004). A unifying pattern underlying
495 these and most other large-scale mammalian radiations over the past 200 million years is that
496 they all involved ecological and locomotory diversification from ancestral lineages of small
497 insectivores (Grossnickle, et al. 2019). The extensive radiation of small terrestrial Eulipotyphla
498 into different adaptive zones, including four independent origins of venom systems in shrews and
499 solenodons (Casewell, et al. 2019) and multiple independent invasions of shrews and moles to
500 semi-aquatic, semi-fossorial, and subterranean environments that occurred on shallow timescales
501 of only a few million years, further demonstrates the high intrinsic evolutionary potential of this
502 Bauplan. Morphological, physiological, and even behavioral convergence have previously been
503 identified within semi-aquatic eulipotyphlan species. For example, *D. moschatus*, *C. cristata*,
504 and *S. palustris* can all detect prey scent while under water via the rapid exhalation and
505 inhalation of air bubbles (Catania 2006; Catania, et al., 2008; Ivlev, et al. 2013), with the latter

506 two species also being characterized by an elevated proton-buffering capacity in muscle
507 (McIntyre, et al. 2002; Gusztak 2008). The results presented here add to this list of
508 convergences, and indicate that semi-aquatic eulipotyphlans have evolved similar Z_{Mb} (and
509 presumably elevated myoglobin concentration) phenotypes via the same selection pressure acting
510 on different sites of the protein and by dissimilar combinations of amino acid substitutions (i.e.,
511 differential gains and losses of cationic and anionic residues, respectively). In other words,
512 molecular adaptation of myoglobin towards life in a semi-aquatic environment is predictable at
513 the protein level but underpinned by unpredictable genotypic evolution. As such, the
514 phylogenomic analysis of myoglobin loci from tissue samples is not only able to provide insights
515 into the lifestyles of rare and recently extinct mammalian species (e.g., museum specimens and
516 subfossil material from the obscure Caribbean nesophontids), but also offers a useful tool to infer
517 past semi-aquatic transitions based on myoglobin primary structure alone.
518

519 **Materials and Methods**

520 *Eulipotyphlan taxon sampling and tree-of-life sequence data collection*

521 Our taxon sampling of eulipotyphlan mammals included 44 talpids, 11 shrews, five
522 erinaceids, and one solenodon (61 specimens encompassing 60 species). Note that this sampling
523 incorporates talpid specimens from five putative 'cryptic lineages' (denoted by 'sp.', 'sp. 1', or
524 'sp. 2' in the figures and tables); for the purpose of this study, each of these genetically distinct
525 lineages are considered independent species. The tissue samples were from various resources
526 (supplementary table S1), with most tissue samples provided by co-authors from China, Japan,
527 Canada, and the USA. Voucher specimens collected by co-authors were deposited in the
528 Kunming Institute of Zoology (KIZ, China), the National Museum of Nature and Science
529 (NMNS, Japan) or kept in personal collections (A.S. and S.I.K.). Additional tissue samples were
530 obtained with permission from the National Museum of Natural History (USNM, USA), the
531 Burke Museum of Natural History and Culture (NWBM, USA), the Field Museum of Natural
532 History (FMNH, USA), and the New Mexico Museum of Natural History (NMMNH, USA).

533 For each specimen, we used a capture hybridization approach (Mason, et al. 2011; Horn
534 2012) to enrich segments of 25 mammalian tree-of-life genes (Meredith, et al. 2011) for
535 phylogenetic analyses. We first downloaded tree-of-life sequences from three eulipotyphlan
536 whole genome sequences available in GenBank (*Erinaceus europaeus*, *Sorex araneus*,

537 *Condylura cristata*), together with 60 bp of 5'- and 3'- flanking sequence for each target. We
538 then aligned each gene segment using MAFFT (Katoh and Standley 2013). The resulting
539 alignments were used to design 120 mer RNA probes (baits) that overlapped by 90 bp (4x tiling),
540 and collapsed any replicates with up to six mismatches (95% similarity) for each segment. For
541 example, if the 120-bp gene fragments from all species were 95% similar with each other, only
542 one probe was designed for this region, otherwise two or more probes were designed to cover the
543 heterogeneity. The probes were synthesized by Arbor Biosciences (Ann Arbor, MI, USA). As a
544 first step in DNA library construction we extracted total DNA from each specimen using a
545 Qiagen DNeasy Blood & Tissue Kit (Qiagen, Canada). The quality and quantity of each DNA
546 sample was measured using a Nanodrop 2000. We then sheared the total DNA into smaller
547 fragments using NEBNext dsDNA Fragmentase (New England Biolabs, Canada), and used this
548 as template to construct DNA libraries using a NEBNext Fast DNA Library Prep Set for Ion
549 Torrent kit (New England Biolabs, Canada). Each sample library contained a unique barcode
550 adapter (NEXTflex DNA Barcodes for Ion Torrent, BIOO Scientific, USA). We selected
551 libraries within the size range of 450-500 bp using a 2% E-gel on an E-Gel Electrophoresis
552 System (Invitrogen, Canada), and re-amplified the size-selected libraries using a NEBNext High-
553 Fidelity 2X PCR Master Mix (New England Biolabs, Canada). Finally, we purified the libraries
554 using Serapure magnetic beads, and measured DNA concentrations using a Qubit 2 Fluorometer
555 (Thermo Fisher Scientific, Canada).
556 We pooled up to four DNA libraries of similar quality and concentrations before
557 hybridization to avoid biased target captures (e.g., baits being used up by one sample).
558 Approximately 500 ng (100 ng to 1000 ng) pooled DNA library was used for each hybridization.
559 We conducted in-solution hybridization using a myBaits® custom target capture kit (Arbor
560 Biosciences, Ann Arbor, MI, USA) following the MYbaits user manual v3.0. The enriched
561 libraries were re-amplified and purified as above. We thereafter measured the DNA
562 concentration using a Qubit flourometer and pooled the enriched libraries for sequencing. The
563 libraries were sequenced using either v318 chips on an Ion Torrent Personal Genome Machine
564 (PGM) or an Ion PI Chip v3 via an Ion Proton Machine.
565
566 *Sequence assembly*

567 Ion Torrent sequencing technology is characterized by higher error rates than Illumina
568 (Jünemann, et al. 2013), and Ion Torrent platforms produce single-end (rather than pair-end)
569 reads. We therefore conducted comprehensive data cleaning and reconciliation procedures, and
570 selected software which could handle single-end sequencing data. The raw data were
571 automatically demultiplexed, trimmed, and converted to FASTQ format on the Torrent Suite
572 v4.0.2 (Thermo Fisher Scientific, Canada) after sequencing. Briefly, we trimmed contaminant
573 (adapters and barcodes) sequences with AlienTrimmer (Criscuolo and Brisse 2013) using
574 conservative parameters (-k 15 -m 5 -l 15 -q 0 -p 0). To remove poor quality data, we used the
575 DynamicTrim function of the software SolexaQA++ v3.1 (Cox, et al. 2010) to trim sequences
576 dynamically and crop the longest contiguous segment for each read. We set the probability value
577 to 0.01 (i.e., one base call error every 100 nucleotides) in this analysis. We removed duplicated
578 and near-duplicated reads for each sample as implemented in ParDRe using all default
579 parameters (Gonzalez-Dominguez and Schmidt 2016). Finally, we conducted data correction
580 using Karect, a multiple sequence alignment-based approach (Allam, et al. 2015), because this
581 software handles substitution, insertion, and deletion errors. The output files of Karect were used
582 for sequence assembly.
583 We de novo assembled the raw sequences for each sample using Abyss v2.0 (Simpson, et al.
584 2009), MRIA v4.0 (Chevreux, et al. 2004), and SPAdes v3.10 (Bankevich, et al. 2012), all of
585 which were designed for short read sequencing data. Abyss is able to use a paired de Bruijn
586 graph instead of a standard de Bruijn graph by specifying a k-mer size (K) and a k-mer pair span
587 (k). We set the K and k to 17 and 33, respectively, and set the maximum number of branches of a
588 bubble to 5 in our analyses. MIRA is based on a Smith-Waterman algorithm. We ran MIRA
589 using specific parameters including bases_per_hash = 31 and minimum_read_length = 35. The
590 SPAdes assembler is also based on a de Bruijn graph, and we set only one k-mer value of 33 for
591 analyses. It is known that merging different draft assemblies (i.e., reconciliation) could improve
592 the assembly quality (Zimin, et al. 2008). We therefore conducted reconciliation using Geneious
593 R11 (https://www.geneious.com). We concatenated the assembled draft contigs generated in
594 three assemblers into a list. We removed contigs shorter than 120 bps, and used the BBMap
595 dedupe function to remove duplicate contigs. We conducted assemblies using the Geneious
596 assembler to group draft contigs with a minimum overlap identity of 96% to a new contig.

597 Finally, all the new contigs and the leftover draft contigs were grouped into a contig list for
598 subsequent analyses.
599
600 *Tree-of-life data analysis*
601 We additionally extracted myoglobin coding sequences and the 25 mammalian tree-of-life
602 gene segments from seven publicly available eulipotyphlan genomes on GenBank using
603 PHYLUCE (Faircloth 2015) and Geneious R11: Indochinese shrew (*Crocidura indochinensis*),
604 gracile shrew-like mole (*Uropsilus nivatus*), Eastern mole (*Scalopus aquaticus*), Hispaniolan
605 solenodon (*Solenodon paradoxus*), European hedgehog (*Erinaceus europaeus*), common shrew
606 (*Sorex araneus*), and the star-nosed mole (*Condylura cristata*). Corresponding sequences from
607 five outgroup taxa were also mined: guinea pig (*Cavia porcellus*), horse (*Equus caballus*), cat
608 (*Felis catus*), pig (*Sus scrofa*) and bat (*Pteropus alecto*). PHYLUCE was originally developed
609 for ultra-conserved elements (UCE). We followed the "harvesting UCE loci from genomes"
610 protocol, but used the tree-of-life reference genes as probes instead of the original UCE probe
611 sets. We extracted genomic regions which were at least 75% similar to the tree-of-life reference
612 sequences. We also mapped the genomes to the tree-of-life references using Geneious with a
613 minimum overlap identity of 75%. These two packages were generally equally efficient at
614 capturing target genes, though in a few cases only one successfully captured the target genes
615 from the genome.
616 We used the above eulipotyphlan myoglobin and tree-of-life gene sequences to generate
617 consensus sequences for the 61 specimens employed for the hybridization capture experiments
618 (supplementary table S2). Briefly, the GenBank sequences were used as reference scaffolds to
619 individually map the Ion Torrent generated reads of each sample using the Geneious "Map to
620 Reference" function, and allowing for a mismatch of 35% per contig. This package conducts
621 iterative mapping and outperforms many other algorithms by higher mapping rates and better
622 consensus accuracy (Kearse, et al. 2012). Approximately 1-4 contigs from each sample were
623 mapped to each gene reference. For the TTN gene segment, whose reference sequence was 4452
624 bp in length, as many as 10 contigs from each sample could be mapped to the reference. In
625 addition, sequences of 19 nuclear gene segments obtained from 21 eulipotyphlan samples
626 collected as part of previous studies (He, et al. 2014; He, et al. 2017), were also used for
627 assemblies as above (and included in the final assemblies). Three shrew species (*Episoriculus*

628 *umbrinus*, *Episoriculus fumidus*, and *Sorex bedfordiae*) for which we obtained myoglobin coding
629 sequences via transcriptome sequencing (see below) were not included in our hybridization
630 capture experiments. We thus downloaded the available tree-of-life genes from these species
631 (*APOB*, *BRCA1*, and *RAG2*) on GenBank and included them in our analysis.
632 The resulting 25 tree-of-life gene segments were aligned separately using MAFFT. We then
633 removed sequences shorter than 247 bp, and estimated gene trees using FastTree v2.1.5 to check
634 for potential non-homologs for each gene. We also checked each alignment by eye, and removed
635 ambiguous regions from the alignment. As segments of two genes (*IRBP*, *PNOC*) failed to
636 hybridize to the majority of the DNA libraries, they were removed from subsequent analysis. The
637 final 23 gene segment alignment (39,414 bp) included sequences from 76 samples
638 (supplementary fig. S11). Because ASTRAL assumes no intra-locus recombination, we also
639 tested recombination events using both RDP and GENECONV methods to examine each gene as
640 implemented in RDP v5.5 (Martin, et al. 2015). When the program detected a signal of
641 recombination, we checked UPGMA trees estimated using the two non-overlapping fragments
642 and also examined the original alignment to see whether the signal was likely due to
643 recombination or other evolutionary processes. We did not observe strong evidence of cross-
644 species recombination in gene alignments (data not shown).
645 We estimated evolutionary relationships using several different approaches. We first
646 constructed a summary-coalescent tree whereby we simultaneously conducted rapid bootstrap
647 analyses and searched for the best scoring maximum likelihood tree using RAXML v8.2
648 (Stamatakis 2014) for each gene alignment, allowing the program to determine the number of
649 bootstraps (-#autoMRE). We used *Cavia porcellus* as the root of the tree employing GTR+ γ
650 during both ML searches and bootstrapping phases, and disabled the BFGS searching algorithm
651 for optimizing branch lengths and GTR parameters (--no-bfgs). We followed Simmons and
652 Kessenich's (2020) recommendation to remove dubiously supported clades and increase accuracy
653 of tree estimation. We used RAxML to estimate SH-like aLRT support values for the best-
654 scoring gene trees (-f j; supplementary table S7). Then we collapsed branches whose SH-like
655 aLRT support values equal zero using Newick utilities and Tree Graph 2 (Junier and Zdobnov
656 2010; Stöver and Müller 2010). Then we used the collapsed trees to estimate the coalescent
657 species tree using ASTRAL III v5.15.0 (Zhang, et al. 2018). Because gene-wise bootstrap could
658 provide more conservative support than site-wise bootstrap analyses (Simmons, et al. 2019), we

659 conducted gene-wise bootstrapping (--gene-only) instead of the typical site-wise bootstrapping.
660 We also allowed the program to explore a larger search space by adding extra bipartitions to the
661 search space (--extraLevel 2).
662 In addition to the summary coalescent analyses, we also constructed a concatenation-based
663 tree using the same dataset. We partitioned the alignment by gene, searched for the best-scoring
664 tree and conducted rapid bootstrapping under the GTR+ γ model using RAxML as described
665 above. We used seven calibrations (supplementary text S1) and BEAST v2.5 (Bouckaert, et al.
666 2014) to estimate divergence times for all 76 samples as well as for the 60 species (55
667 eulipotyphlans and 5 outgroup species) for which complete myoglobin coding sequences were
668 obtained. For this analysis, we first partitioned the alignment by gene and used the bModelTest
669 package of BEAST2 to estimate the most appropriate substitution model for each gene
670 (Bouckaert and Drummond 2017). We used a relaxed clock model with lognormal distribution
671 for estimating the branch lengths, a birth-death model for the prior of the tree, and ran the
672 analysis for 100 million generations. We used *Cavia porcellus* as the outgroup to Laurasiatheria,
673 and also fixed the relationships of the other four outgroup species, because a biased sampling
674 toward the ingroup (i.e., Eulipotyphla) may lead to an inaccurate estimation of outgroup
675 relationships (Springer, et al. 2018). Secondly, we used the BModelAnalyzer package of
676 BEAST2 to determine the best models for each gene based on the results of the bModelTest
677 (supplementary table S6; Barido-Sottani, et al. 2018). We fixed the models of evolution based on
678 the results of BModelAnalyzer and re-ran BEAST using the same parameters described above.
679 Finally, we employed a multispecies coalescent model (*BEAST; Heled and Drummond
680 2010) as implemented in BEAST v2.5. We grouped the samples by species. We used linear and
681 constant root as the prior for the population model. Substitution model, tree model, and
682 calibrations were set as above. All gene trees and species trees are given in supplementary table
683 S8.
684 To examine whether alternative evolutionary hypotheses of life histories (e.g., a single
685 origination of fully fossorial lifestyle within Talpidae) could be statistically rejected, we
686 performed Shimodaira–Hasegawa (SH) tests. For these analyses, we constrained the
687 monophyletic relationships of: i) fully fossorial Talpini and Scalopini moles, ii) semi-aquatic
688 desmans and the star-nosed mole, and iii) semi-aquatic nectogaline shrew genera (*Chimarrogale*,
689 *Nectogale*, and *Neomys*), one at a time and estimated the maximum likelihood concatenation

690 trees using RAxML as described above (supplementary table S3). Then we computed the log
691 likelihood between the best scoring maximum likelihood tree and the constrained alternative
692 phylogenies as implemented in RAxML (-f H).
693
694 *Myoglobin collection and analysis*
695 We used four strategies to obtain myoglobin coding sequences (supplementary table S1). We
696 first extracted available eulipotyphlan myoglobin mRNA and gene sequences from GenBank.
697 The three coding exons were individually used both as templates for capture hybridization probe
698 design (see above) and to map the hybridization contigs/generate consensus sequences for each
699 exon. The 5'- and 3'- ends of introns were confirmed based on the GT-AG splice site rule. Of the
700 61 samples that we used for hybridization capture, complete coding sequence was obtained for
701 27 samples, partial myoglobin sequences were obtained for 30 samples, and no sequence
702 obtained for 4 samples (supplementary table S1).
703 To cross-validate the results of our hybridization experiments, fill sequencing gaps, and
704 extend our taxon sampling, we also PCR amplified and Sanger sequenced whole myoglobin
705 exons from existing DNA samples and/or employed transcriptome sequencing on additional
706 eulipotyphlan specimens. For the latter, we collected heart and lung samples from five shrews,
707 one shrew-like mole, and one gymnure **(**supplementary table S1). Tissues were preserved in
708 RNA*later* (Qiagen, China), and stored at -80C. Total RNA was extracted using a RNeasy Mini
709 kit (Qiagen, China), and mRNA subsequently enriched using immobilized oligo(dT). mRNA was
710 sheared and reverse transcribed to DNA. The cDNA libraries were purified and re-amplified
711 using PCR for de novo sequencing using a HiSeq X™ Ten Sequencing System. Approximately 6
712 Gb data were obtained for each sample. Experiments and sequencing were conducted by
713 BioMarker Co. (Beijing, China). We used FastQC v0.11.5 (Andrews 2010) to access sequence
714 quality, and trimmed adapter sequences using Trimmomatic v0.39 (Bolger, et al. 2014). We
715 conducted *de no* assembly using Trinity v2.4 with default parameters (Grabherr, et al. 2011).
716 Finally, primers for PCR were designed for conserved exon flanking regions from available
717 eulipotyphlan genomes, hybridization capture, and mRNA sequences, and were used for both
718 PCR and Sanger sequencing (supplementary table S9). These procedures resulted in complete
719 coding sequences being obtained for 55 eulipotyphlan species (supplementary table S4). We
720 estimated the myoglobin gene trees utilizing both nucleotide and amino acid sequences as

721 implemented in RAxML. We used Dayhoff+ γ model (Dayhoff, et al. 1978) for the amino acid
722 gene tree estimation using the same settings described above.
723
724 *Ancestral sequence reconstruction and homology modelling*
725 We estimated ancestral myoglobin sequences for each node of a 60 species phylogeny that
726 utilized both DNA sequences and amino acids. Prior to analysis, the start (methionine) and stop
727 codons were removed from the alignment. As in our previous study (Mirceta, et al. 2013), we
728 performed maximum likelihood ancestral amino acid sequence reconstruction as implemented in
729 MEGA (Kumar, et al. 2018) using the Dayhoff+ γ model that was obtained as the best-fitting
730 substitution model using the model test facility in MEGA-X. Prior to conducting the codon-
731 based analysis, we removed codons corresponding to residue position 121 (which was absent for
732 5 of the 60 species; supplementary table S4). We then used the PAML package CodeML (Yang
733 2007) as implemented in EasyCodeML (Gao, et al. 2019), and compared codon substitution
734 models (site models) including M0, M1a, M2a, M3, M7, M8 and M8a using likelihood-ratio
735 tests. We relied on the model with the highest likelihood (M8a). Because PAML does not take
736 account of insertion/deletion events (indels), and instead treats gaps as missing data, the ancestral
737 states of the gapped codon position 121 was reconstructed separately using a likelihood-based
738 mixture model as implemented in FastML (Ashkenazy, et al. 2012).
739 To assess the three-dimensional location and any secondary or tertiary structural implications
740 of amino acid replacements or insertions/deletions, we used the fully automated homology
741 modelling facilities of the SWISS-MODEL server (Waterhouse, et al. 2018) to build protein
742 structural models from the reconstructed ancestral primary structure of myoglobin in the last
743 common eulipotyphlan ancestor and from the primary structure of the sequenced myoglobins of
744 one species of each of the five semiaquatic lineages. Implications of the gapped position 121 on
745 the tertiary structure of myoglobin in the Russian desman compared to the last eulipotyphlan
746 ancestor were visualised in PyMol (The PyMOL Molecular Graphics System, Version 2.1.1,
747 Schrödinger, LLC).
748
749 *Calculation of myoglobin net surface charge and electrophoretic mobility*
750 We calculated Z_{Mb} as the sum of the charge of all ionizable groups in myoglobin at pH = 6.5
751 by modelling Mb primary structures onto the tertiary structure and using published, conserved,

752 site-specific ionization constants (McLellan 1984; Mirceta, et al. 2013). The reliability of
753 modelled Z_{Mb} values was assessed by determining the electrophoretic mobility of native
754 myoglobin bands at the same pH in muscle extracts of representative eulipotyphlan species and
755 the grey seal, *Halichoerus grypus*, as an example of a marine mammal. Approximately 0.2 g of
756 skeletal or cardiac muscle tissue from selected species, freed from any obvious fat or connective
757 tissue remnants and rinsed with homogenisation buffer to move any remaining blood, was
758 homogenized in 5 volumes of ice-cold 0.2 M MES buffer [2-(N-morpholino)ethanesulfonic acid]
759 adjusted to pH 6.5, using an Ultra-turrax T25 homogeniser for 10 s at first 9,500 rpm and then 3
760 times at 13,500 rpm, leaving samples to cool down between steps for 1 min on ice to avoid heat
761 denaturation of proteins. The homogenized muscle extracts were then centrifuged at 10,500 g (20
762 min at 4°C) and the supernatants stored at -80°C until further use. Electrophoretic mobility of
763 thawed muscle extracts was assessed in 9% polyacrylamide gels containing 0.3 M MES buffer
764 pH 6.5, using a Bio-Rad Mini-PROTEAN II gel system with 0.2 M MES pH 6.5 as the running
765 buffer at 100 V and room temperature for a minimum of 3 hours. Native myoglobin bands were
766 identified by their distinct red-brown colour before general protein staining with EZBlue™
767 (G104, Sigma-Aldrich). Electrophoretic mobility was assessed on digital gel images and
768 expressed as distance travelled relative to the grey seal myoglobin, which was used as a standard
769 of a marine mammal myoglobin with high net surface charge (Mirceta, et al. 2013). The
770 correlation between measured relative electrophoretic mobility and modelled Z_{Mb} values was
771 assessed using Phylogenetic Generalized Least Squares (PGLS) analysis using the CAPER
772 package (Orme, et al. 2013) as implemented in R v3.6.1 and the tree from the BEAST analysis in
773 Fig. 2. Because of low sample size, the parameter lambda was not estimated from the data but
774 fixed at a value of 1.0.
775
776 *Lifestyle correlation analysis and ancestral lifestyle reconstruction*
777 We analyzed the relationship between lifestyles and Z_{Mb} based on a threshold model
778 (Felsenstein 2012) using the phytools function threshBayes (Revell 2012) as implemented in R
779 v3.6. The threshold model hypothesizes that each lifestyle is determined by an underlying,
780 unobserved continuous trait (i.e., liability). We first categorized the 55 eulipotyphlan species for
781 which Z_{Mb} was calculated as either semi-aquatic (including the semi-aquatic/fossorial star-nosed
782 mole in this category) or non-aquatic based on the habits described in Burgin and He (2018). We

783 ran Markov chain Monte Carlo (MCMC) for 10^7 generations, sampling every 500 generations,
784 and discarded the first 20% generations as burn-in. We plotted the posterior sample for the
785 correlation to examine whether analyses reached a stationary state. We also estimated the
786 correlation between Z_{Mb} and full fossoriality (i.e., fossorial species versus non-fossorial species),
787 as well as that between Z_{Mb} and 'digging' (a category that included both fossorial and semi-
788 fossorial species) habits using the same approach. Finally, we also created subsets of our dataset
789 to enable comparisons between only two ecomorphotypes, with the following four threshBayes
790 analyses conducted: terrestrial Z_{Mb} versus semi-aquatic Z_{Mb}, terrestrial Z_{Mb} versus fully fossorial
791 Z_{Mb}, terrestrial Z_{Mb} versus semi-fossorial/fossorial Z_{Mb}, fully fossorial Z_{Mb} versus semi-aquatic
792 Z_{Mb}.
793 We estimated the ancestral lifestyle using a maximum parsimony and a threshold model
794 based on the 76 species time-calibrated concatenated gene tree (fig. 2) and the *BEAST
795 coalescent species tree (fig. S3). We categorized the species into non-aquatic or semi-aquatic as
796 above. We used the R package castor to reconstruct ancestral lifestyles using maximum
797 parsimony (Louca and Doebeli 2017), treating the transition cost between non-aquatic and semi-
798 aquatic equally. We then used the R package phytools to estimate ancestral states using a
799 threshold model (Revell 2014). We ran MCMC for 1 million generations, sampling every 1000
800 generations, and discarded the first 20% generations as burn-in. We performed the threshold
801 analyses using either a Brownian motion or lambda model for estimating the liability, and
802 compared the results based on deviance information criterion (DIC). We selected the result of the
803 lambda model because it outperformed the Brownian motion model (ΔDIC =99, data not
804 shown).
805
806 **Acknowledgements**
807 We thank curators and staff from the Smithsonian Institute, National Museum of Natural History
808 (K.M. Helgen), Field Museum (L. Heaney), Burke Museum (B. Sharon) and New Mexico
809 Museum of Natural History (B. Oh), National Museum of Natural Science (Y. Chen), Kunming
810 Institute of Zoology (S. Li and X-L. Jiang) for approving our proposal for the use of tissue
811 samples, or conducting destructive sampling on museum skin specimens. We are grateful to K.
812 Wareing and S. Mirceta, University of Liverpool for muscle samples of the European mole and
813 hedgehog, and for help with native PAGE of muscle extracts, respectively. We thank M. Docker,

University of Manitoba, for use of the Ion torrent PGM, and Z.-L. Ding, Kunming Institute of Zoology, for performing Ion Torrent Proton sequencing. We appreciate Ailaoshan Station for Subtropical Forest Ecosystem Studies, Xishuangbanna Tropical Botanical Garden, Chinese Academy of Sciences for allowing us to conduct fieldwork. We also thank Y-H. Sun and S-Y. Liu for supporting our fieldwork in Gansu and Sichuan provinces, and Yuan and L-K. Lin for assistance with fieldwork in Taiwan, and Z. Liu for providing tissue samples. We are grateful to Y-P. Wang (Soochow university) for conducting sequencing. Finally, we are grateful to Umi Matsushita for painting eulipotyphlan species for our figures. This work was supported by the National Natural Science Foundation of China (31970389, 31301869 to K. H.), the National Science Foundation (NSF DEB-1457735 to M.S.S), and by the University of Manitoba Research Grants Program (41342), National Sciences and Engineering Research Council of Canada Discovery (RGPIN/238838-2011; RGPIN/6562-2016) and Discovery Accelerator Supplement (RGPIN/412336-2011) grants to K.L.C.

Competing interests

The authors declare that they have no conflict of interest.

Data Availability Statement

The newly obtained myoglobin sequences were deposited to GenBank under accession numbers MW456061 to MW456069 and MW473727- MW473769, and sequence alignments per gene were deposited to Dryad Digital Repository at https://doi.org/10.5061/dryad.brv15dv7q.

Supplementary files

Fig. S1. The full RAxML species tree constructed from 71 eulipotyphlan specimens. Bootstrap supports are given next to internal nodes.

Fig. S2. The full ASTRAL-III coalescent species tree constructed from 71 eulipotyphlan specimens. Bootstrap supports are given next to internal nodes.

Fig. S3. The full *BEAST coalescent species tree constructed from 71 eulipotyphlan specimens. Posterior probabilities are given next to internal nodes.

845

846 **Fig. S4.** Location of charge-changing amino acid substitutions in the oxygen-storing protein
847 myoglobin of four semi-aquatic species of moles and shrews in the mammalian insectivore order
848 Eulipotyphla. Three-dimensional structures were obtained by homology modelling using the
849 SWISS-MODEL server (Waterhouse, et al. 2018) and the amino acid sequences of (A) *Sorex*
850 *palustris*, (B) *Neomys fodiens*, (C) *Nectogale elegans*, and (D) *Condylura cristata*. The three-
851 dimensional myoglobin structure of the last common ancestor of the group was also modelled
852 based on results of an amino acid sequence reconstruction (see text for details). Ancestral (left)
853 and derived (right) states of charge-changing amino acid replacements are circled and indicated
854 with positional number and one-letter amino acid code. Blue and red color indicate amino acids
855 with positively (H, His; K, Lys; R, Arg) and negatively charged amino acid side chains (D, Asp;
856 E, Glu). White double arrows indicate salt bridges that are affected by charge-changing
857 substitutions. Image views between panels A-D have been rotated to maximally visualise
858 lineage-specific replacements.

859

860 **Fig. S5**: Structural model of myoglobin in (A) the last common ancestor of Eulipotyphla and (B)
861 the semi-aquatic Russian desman (*Desmana moschata*). Structures were obtained by homology
862 modelling using the SWISS-MODEL server (Waterhouse, et al. 2018) and the primary structures
863 of myoglobin obtained by conceptual translation of the here determined nucleotide sequence
864 (Russian desman) or by ancestral amino acid sequence reconstruction (see text for details).
865 Structures were visualised in PyMOL version 2.1.1. Note that the gap at position 121 in the GH-
866 loop of the tertiary structure (circled in red) of the Russian desman in (B) appears to exert
867 negligible effect on the tertiary structure of the protein.

868

869 **Fig. S6**. Time-calibrated tree of 55 eulipotyphlan species for which complete myoglobin coding
870 sequences were determined (left). Horizontal bars on the right indicate the calculated Z_{Mb} for
871 each species, which are color coded according to species lifestyle.

872

873 **Fig. S7**. Maximum likelihood ancestral (A) amino acid-based and (B) codon-based sequence
874 reconstruction and net surface charge (Z_{Mb}) calculation of eulipotyphlan myoglobin mapped onto
875 the time calibrated phylogeny of fig. 2. Results of a separate amino-acid based reconstruction

876 based on the *BEAST species tree reconstruction (supplementary fig. S3) is provided in (C) with
877 *Episoriculus fumidus* as the species with an alternative position compared to *A* and *B* in red font.
878 Only charge-changing substitutions are shown, with blue and red font in text boxes indicating
879 charge increasing and charge decreasing substitutions, respectively. Absolute charge-changing
880 substitutions >1.0 are underlined, whereas those <0.10 are shown in brackets. Yellow
881 highlighting indicates amino acids reconstructed with $p < 0.95$ for which less likely but
882 differentially charged amino acids have been reconstructed with $p > 0.05$ (see text for details).
883 Reconstructed net charges are shown at nodes, though have been omitted in some cases if they
884 were identical to values on the preceding node, for clarity. Terminal species considered semi-
885 aquatic are indicated by blue font, with Z_{Mb} values >+2.0 also given in blue font.
886
887 **Fig. S8**. Comparisons of the Bayesian concatenation species tree estimated using the tree-of-life
888 genes with myoglobin RAxML gene trees estimated using (A) nucleotide and (B) amino-acid
889 sequences. Bootstrap supports are given next to internal nodes on the myoglobin gene trees. Only
890 Bootstrap supports higher than 70 are shown.
891
892 **Fig. S9**. Ancestral reconstruction of semi-aquatic lifestyles (blue filled circles) within
893 Eulipotyphla based on both the (A) RAxML concatenation gene tree and (B) the *BEAST
894 species tree. For each tree, reconstructions were constructed using either a maximum parsimony
895 (left) or a threshold (right) model. Note that *Episoriculus fumidus* (red font in *B*) was supported
896 at different positions on the two trees.
897
898 **Fig. S10**. Posterior density distribution of the correlation coefficient (r) between (A) semi-
899 aquatic, (B) fully fossorial, and (C) digging (to include fossorial and semi-fossorial lifestyles)
900 and Z_{Mb} as estimated using the threshBayes function. Figures D-G show the results based on the
901 analyses of subsets of samples that included only: (D) terrestrial and semi-aquatic species, (E)
902 terrestrial and fully fossorial species, (F) terrestrial and digging species, or (G) semi-aquatic and
903 fully fossorial species. The solid red line indicates the grand mean, the green dashed lines
904 indicate the mean of the 80% confidence intervals, and the grey dashed lines indicate the mean
905 of the 95% confidence intervals.
906

907 **Fig. S11**. A heatmap showing tree-of-life genes included for each of the 76 samples used for
908 phylogenetic analyses. The ultrametric tree is the BEAST concatenation gene tree and a blank
909 block indicates the gene is missing in the final dataset.
910
911 **Table S1**. Sample information of specimens used in this study.
912
913 **Table S2**. Hybridization capture results of tree-of-life gene segments from 61 eulipotyphlan
914 DNA libraries. Numbers in each column represent total base pairs captured; NA: no data.
915
916 **Table S3.** Result of likelihood-based Shimodaira–Hasegawa test to compare the best scoring
917 RAxML concatenated gene tree and alternative evolutionary hypotheses.
918
919 **Table S4**. Myoglobin amino acid alignment used for modeling myoglobin net surface charge
920 (Z_{Mb}) and ancestral sequence reconstructions. Myoglobin helices A to H are highlighted in
921 yellow, with amino acid positions and helical notations indicated above and below the graphic,
922 respectively. Internal amino acid residue positions are shaded in light grey, while deleted
923 residues are indicated by a dash mark. Strongly anionic residues (D [Asp] and E [Glu]) are
924 shaded in red, with strongly (K [Lys] and R [Arg]) and weakly (H [His]) cationic residues
925 shaded in dark and light green, respectively.
926
927 **Table S5**. Charge increasing (blue font) and decreasing (red font) residue substitutions
928 reconstructed for semi-aquatic eulipotyphlan branches.
929
930 **Table S6**. Evolutionary models estimated using bModelTest in BEAST, and used for BEAST
931 and *BEAST analyses.
932
933 **Table S7.** RAxML best scoring gene trees used for ASTRAL-III coalescent analysis before and
934 after collapsing 0% Shimodaira–Hasegawa (SH) scores in Newick format.
935

936 **Table S8.** The best scoring concatenation species trees estimated using BEAST and RAxML,
937 and the best species coalescence trees estimated using ASTRAL-III and *BEAST, in Newick
938 format.
939
940 **Table S9.** Primers used to amplify and sequence the protein coding exons of myoglobin.
941
942 **Text S1**. Calibrations used for estimating divergence times in the BEAST analyses.
943
944

945 ## References

946 Allam A, Kalnis P, Solovyev V. 2015. Karect: accurate correction of substitution, insertion and
947 deletion errors for next-generation sequencing data. *Bioinformatics* 31:3421–3428.
948
949 Andrews S. 2010. FastQC: a quality control tool for high throughput sequence data. In:
950 Babraham Bioinformatics, Babraham Institute, Cambridge, United Kingdom.
951
952 Ashkenazy H, Penn O, Doron-Faigenboim A, Cohen O, Cannarozzi G, Zomer O, Pupko T. 2012.
953 FastML: a web server for probabilistic reconstruction of ancestral sequences. *Nucleic Acids Res.*
954 40:W580–W584.
955
956 Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko
957 SI, Pham S, Prjibelski AD. 2012. SPAdes: a new genome assembly algorithm and its
958 applications to single-cell sequencing. *J Comput Biol.* 19:455–477.
959
960 Barido-Sottani J, Boskova V, Plessis LD, Kuhnert D, Magnus C, Mitov V, Muller NF, PecErska
961 J, Rasmussen DA, Zhang C, et al. 2018. Taming the BEAST-A community teaching material
962 resource for BEAST 2. *Syst Biol.* 67:170–174.
963
964 Berenbrink M. 2020. The role of myoglobin in the evolution of mammalian diving capacity –
965 The August Krogh principle applied in molecular and evolutionary physiology. *Comp Biochem*
966 *Physiol A Mol Integr Physiol.* 252:110843.

967
968 Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence
969 data. *Bioinformatics* 30:2114–2120.
970
971 Bosshard HR, Marti DN, Jelesarov I. 2004. Protein stabilization by salt bridges: concepts,
972 experimental approaches and clarification of some misunderstandings. *J Mol Recognit.* 17:1–16.
973
974 Bouckaert R, Heled J, Kühnert D, Vaughan T, Wu C-H, Xie D, Suchard MA, Rambaut A,
975 Drummond AJ. 2014. BEAST 2: a software platform for Bayesian evolutionary analysis. *PLoS*
976 *Comput Biol.* 10:e1003537.
977
978 Bouckaert RR, Drummond AJ. 2017. bModelTest: Bayesian phylogenetic site model averaging
979 and model comparison. *BMC Evol Biol.* 17:42.
980
981 Brace S, Thomas JA, Dalen L, Burger J, MacPhee RDE, Barnes I, Turvey ST. 2016.
982 Evolutionary history of the Nesophontidae, the last unplaced recent mammal family. *Mol Biol*
983 *Evol.* 33:3095–3103.
984
985 Burgin C, He K. 2018. Family Soricidae (Shrews). In: Wilson DE, Mittermeier RA, editors.
986 Handbook of The Mammals of the World. Vol. 8. Insectivores, Sloths and Colugos. Barcelona:
987 Lynx Edicions. p. 332–551.
988
989 Burgin CJ, Colella JP, Kahn PL, Upham NS. 2018. How many species of mammals are there? *J*
990 *Mammal.* 99:1–14.
991
992 Butler PM. 1998. Fossil history of shrews in Africa. In: Wojcik JM, Wolsan M, editors.
993 Evolution of Shrews. Bialowieza: Mammal Research Institute. p. 121–132.
994
995 Casewell NR, Petras D, Card DC, Suranse V, Mychajliw AM, Richards D, Koludarov I,
996 Albulescu LO, Slagboom J, Hempel BF, Ngum NM. 2019. Solenodon genome reveals

997 convergent evolution of venom in eulipotyphlan mammals. *Proc Nat Acad Sci U S A.*
998 116:25745–25755.
999
1000 Campbell B. 1939. The shoulder anatomy of the moles. A study in phylogeny and adaptation.
1001 *Am J Anat.* 64:1–39.
1002
1003 Catania KC. 2006. Olfaction - Underwater 'sniffing' by semi-aquatic mammals. *Nature*
1004 444:1024–1025.
1005
1006 Catania KC, Hare JF, Campbell KL. 2008. Water shrews detect movement, shape, and smell to
1007 find prey underwater. *Proc Nat Acad Sci U S A.* 105:571–576.
1008
1009 Chevreux B, Pfisterer T, Drescher B, Driesel AJ, Muller WEG, Wetter T, Suhai S. 2004. Using
1010 the miraEST assembler for reliable and automated mRNA transcript assembly and SNP detection
1011 in sequenced ESTs. *Genome Res.* 14:1147-1159.
1012
1013 Churchfield S. 1990. The Natural History of Shrews: Cornell University Press.
1014
1015 Cox MP, Peterson DA, Biggs PJ. 2010. SolexaQA: At-a-glance quality assessment of Illumina
1016 second-generation sequencing data. *BMC Bioinformatics* 11:485.
1017
1018 Criscuolo A, Brisse S. 2013. AlienTrimmer: a tool to quickly and accurately trim off multiple
1019 short contaminant sequences from high-throughput sequencing reads. *Genomics* 102:500–506.
1020
1021 Dayhoff M, Schwartz R, Orcutt B. 1978. A model of evolutionary change in proteins. In. Atlas
1022 of protein sequence and structure: National Biomedical Research Foundation Silver Spring MD.
1023 p. 345–352.
1024
1025 Emerling CA, Springer MS. 2014. Eyes underground: regression of visual protein networks in
1026 subterranean mammals. *Mol Phylogenet Evol.* 78:260–270.
1027

1028 Enoki Y, Ohga Y, Ishidate H, Morimoto T. 2008. Primary structure of myoglobins from 31
1029 species of birds. *Comp Biochem Physiol B Biochem Mol Biol.* 149:11–21.
1030
1031 Esselstyn JA, Oliveros CH, Swanson MT, Faircloth BC. 2017. Investigating difficult nodes in
1032 the placental mammal tree with expanded taxon sampling and thousands of ultraconserved
1033 elements. *Genome Biol Evol.* 9:2308–2321.
1034
1035 Faircloth BC. 2015. PHYLUCE is a software package for the analysis of conserved genomic
1036 loci. *Bioinformatics* 32:786–788.
1037
1038 Felsenstein J. 2012. A comparative method for both discrete and continuous characters using the
1039 threshold model. *Am Nat.* 179:145–156.
1040
1041 Gao F, Chen C, Arab DA, Du Z, He Y, Ho SYW. 2019. EasyCodeML: A visual tool for analysis
1042 of selection using CodeML. *Ecol Evol.* 9:3891–3898.
1043
1044 Gonzalez-Dominguez J, Schmidt B. 2016. ParDRe: faster parallel duplicated reads removal tool
1045 for sequencing studies. *Bioinformatics* 32:1562–1564.
1046
1047 Grabherr M, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, Adiconis X, Fan L,
1048 Raychowdhury R, Zeng QJNB. 2011. Full-length transcriptome assembly from RNA-Seq data
1049 without a reference genome. *Nat Protoc.* 29:644-652.
1050
1051 Gorman ML, Stone RD. 1990. The Natural History of Moles: Cornell University Press.
1052
1053 Grand T, Gould E, Montali R. 1998. Structure of the proboscis and rays of the star-nosed mole,
1054 *Condylura cristata*. *J Mammal.* 79:492–501.
1055
1056 Grossnickle DM, Smith SM, Wilson GP. 2019. Untangling the multiple ecological radiations of
1057 early mammals. *Trends Ecol Evol.* 34:936-949.
1058

1059 Gusztak RW. 2008. Dive performance and aquatic thermoregulation of the world's smallest
1060 mammalian diver, the American water shrew (*Sorex palustris*). M.Sc. thesis. University of
1061 Manitoba.
1062
1063 Gusztak RW, MacArthur RA, Campbell KL. 2005. Bioenergetics and thermal physiology of
1064 American water shrews (*Sorex palustris*). *J Comp Physiol B.* 175:87–95.
1065
1066 Hahn MW, Nakhleh L. 2016. Irrational exuberance for resolved species trees. *Evolution* 70:7-17.
1067 He K, Chen J-H, Gould GC, Yamaguchi N, Ai H-S, Wang Y-X, Zhang Y-P, Jiang X-L. 2012.
1068 An estimation of Erinaceidae phylogeny: A combined analysis approach. *PLoS ONE* 7:e39304.
1069
1070 He K, Li YJ, Brandley MC, Lin LK, Wang YX, Zhang YP, Jiang XL. 2010. A multi-locus
1071 phylogeny of Nectogalini shrews and influences of the paleoclimate on speciation and evolution.
1072 *Mol Phylogenet Evol.* 56:734–746.
1073
1074 He K, Shinohara A, Helgen KM, Springer MS, Jiang X-L, Campbell KL. 2017. Talpid mole
1075 phylogeny unites shrew moles and illuminates overlooked cryptic species diversity. *Mol Biol*
1076 *Evol.* 34:78–87.
1077
1078 He K, Shinohara A, Jiang X-L, Campbell KL. 2014. Multilocus phylogeny of talpine moles
1079 (Talpini, Talpidae, Eulipotyphla) and its implications for systematics. *Mol Phylogenet Evol.*
1080 70:513–521.
1081
1082 Heled J, Drummond AJ. 2010. Bayesian inference of species trees from multilocus data. *Mol*
1083 *Biol Evol.* 27:570–580.
1084 Hickman GC. 1984. Swimming ability of talpid moles, with particular reference to the semi-
1085 aquatic *Condylura cristata*. *Mammalia* 48:505–514.
1086

1087 Hooker JJ. 2016. Skeletal adaptations and phylogeny of the oldest mole Eotalpa (Talpidae,
1088 Lipotyphla, Mammalia) from the UK Eocene: the beginning of fossoriality in moles.
1089 *Palaeontology* 59:195–216.
1090
1091 Horn S. 2012. Target enrichment via DNA hybridization capture. In: Shapiro B, Hofreiter M,
1092 editors. Ancient DNA: Springer. p. 177–188.
1093
1094 Hutterer R. 1985. Anatomical adaptations of shrews. *Mamm Rev.* 15:43–55.
1095
1096 Isogai Y, Imamura H, Nakae S, Sumi T, Takahashi K-i, Nakagawa T, Tsuneshige A, Shirai T.
1097 2018. Tracing whale myoglobin evolution by resurrecting ancient proteins. *Sci Rep.* 8:16883.
1098
1099 Ivlev YF, Kuznetsov A, Rutovskaya M. 2010. Preliminary data on the swimming kinematics of
1100 the Russian desman (*Desmana moschata* L.). *Dokl Biol Sci.* 431:144–148.
1101 Ivlev YF, Rutovskaya MV, Luchkina OS. (Ivlev2013 co-authors). 2013. The use of olfaction by
1102 the Russian desman (*Desmana moschata* L.) during underwater swimming. *Dokl Biol Sci.*
1103 452:280–283.
1104 Junemann S, Sedlazeck FJ, Prior K, Albersmeier A, John U, Kalinowski J, Mellmann A,
1105 Goesmann A, von Haeseler A, Stoye J, Harmsen D. 2013. Updating benchtop sequencing
1106 performance comparison. *Nat Biotechnol.* 31:294–296.
1107
1108 Junier T, Zdobnov EM. 2010. The Newick utilities: high-throughput phylogenetic tree
1109 processing in the UNIX shell. *Bioinformatics* 26:1669–1670.
1110
1111 Katoh K, Standley DM. 2013. MAFFT multiple sequence alignment software version 7:
1112 improvements in performance and usability. *Mol Biol Evol.* 30:772–780.
1113
1114 Kearse M, Sturrock S, Meintjes P. 2012. The Geneious 6.0. 3 read mapper. Auckland, New
1115 Zealand: Biomatters, Ltd.
1116

1117 Kumar S, Stecher G, Li M, Knyaz C, Tamura K. 2018. MEGA X: molecular evolutionary
1118 genetics analysis across computing platforms. *Mol Biol Evol.* 35:1547–1549.
1119
1120 Lestyk KC, Folkow LP, Blix AS, Hammill MO, Burns JM. 2009. Development of myoglobin
1121 concentration and acid buffering capacity in harp (*Pagophilus groenlandicus*) and hooded
1122 (*Cystophora cristata*) seals from birth to maturity. *J Comp Physiol B* 179:985–996.
1123
1124 Louca S, Doebeli M. 2017. Efficient comparative phylogenetics on large trees. *Bioinformatics*
1125 34:1053–1055.
1126
1127 Madsen O, Scally M, Douady CJ, Kao DJ, DeBry RW, Adkins R, Amrine HM, Stanhope MJ, de
1128 Jong WW, Springer MS. 2001. Parallel adaptive radiations in two major clades of placental
1129 mammals. *Nature* 409:610–614.
1130
1131 Martin DP, Murrell B, Golden M, Khoosal A, Muhire B. 2015. RDP4: Detection and analysis of
1132 recombination patterns in virus genomes. *Virus Evol* 1:vev003.
1133
1134 MacPhee RDE, Flemming C, Lunde DP. 1999. "Last occurrence" of the Antillean insectivoran
1135 Nesophontes: new radiometric dates and their interpretation. *Am Mus Novit.* 3261:1–19.
1136
1137 Mason VC, Li G, Helgen KM, Murphy WJ. 2011. Efficient cross-species capture hybridization
1138 and next-generation sequencing of mitochondrial genomes from noninvasively sampled museum
1139 specimens. *Genome Res.* 21:1695–1704.
1140
1141 McIntyre IW, Campbell KL, MacArthur RA. 2002. Body oxygen stores, aerobic dive limits and
1142 diving behaviour of the star-nosed mole (*Condylura cristata*) and comparisons with non-aquatic
1143 talpids. *J Exp Biol.* 205:45–54.
1144
1145 McLellan T. 1984. Molecular charge and electrophoretic mobility in cetacean myoglobins of
1146 known sequence. *Biochem Genet.* 22:181–200.
1147

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

Mirceta S, Signore AV, Burns JM, Cossins AR, Campbell KL, Berenbrink M. 2013. Evolution of mammalian diving capacity traced by myoglobin net surface charge. *Science* 340:1234192.

Mittermeier RA, Wilson DE, Rylands AB. 2018. Handbook of the Mammals of the World: Insectivores, Sloths and Colugos. Barcelona Lynx Edicions.

Motokawa M. 2004. Phylogenetic relationships within the family Talpidae (Mammalia: Insectivora). *J Zool (Lond.)* 263:147–157.

Mulugeta E, Wassenaar E, Sleddens-Linkels E, van Ijcken WFJ, Heard E, Grootegoed JA, Just W, Gribnau J, Baarends WM. 2016. Genomes of *Ellobius* species provide insight into the evolutionary dynamics of mammalian sex chromosomes. *Genome Res.* 26:1202–1210.

Noren SR, Williams TM. 2000. Body size and skeletal muscle myoglobin of cetaceans: adaptations for maximizing dive duration. *Comp Biochem Physiol A Mol Integr Physiol.* 126:181–191.

Nowak RM. 1999. Walker's Mammals of the World. Baltimore: Johns Hopkins University Press.

Orme D, Freckleton R, Thomas G, Petzoldt T, 2013. The caper package: comparative analysis of phylogenetics and evolution in R. *R package version* 5:1–36.

Piras P, Sansalone G, Teresi L, Kotsakis T, Colangelo P, Loy A. 2012. Testing convergent and parallel adaptations in talpids humeral mechanical performance by means of geometric morphometrics and finite element analysis. *J Morphol.* 273:696–711.

Ponganis PJ, Meir JU, Williams CL. 2011. In pursuit of Irving and Scholander: a review of oxygen store management in seals and penguins. *J Exp Biol.* 214:3325–3339.

Reed CA. 1951. Locomotion and appendicular anatomy in three soricoid insectivores. *Am Midl Nat.* 45:513–671.

Revell LJ. 2014. Ancestral character estimation under the threshold model from quantitative genetics. *Evolution* 68:743–759.

Revell LJ. 2012. phytools: an R package for phylogenetic comparative biology (and other things). *Methods Ecol Evol.* 3:217–223.

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

Romero-Herrera A, Lehmann H, Joysey K, Friday A. 1973. Molecular evolution of myoglobin and the fossil record: a phylogenetic synthesis. *Nature* 246:389–395.

Romero-Herrera A, Lehmann H, Joysey K, Friday A. 1978. On the evolution of myoglobin. *Philos Trans R Soc Lond B Biol Sci.* 283:61–163.

Samuel, P. P., Smith, L. P., Phillips, G. N., Jr., and Olson, J. S. (2015) Apoglobin stability Is the major factor governing both cell-free and in vivo expression of holomyoglobin. *J. Biol. Chem.* 290: 23479-23495.

Sanchez-Villagra MR, Horovitz I, Motokawa M. 2006. A comprehensive morphological analysis of talpid moles (Mammalia) phylogenetic relationships. *Cladistics* 22:5–88.

Sansalone G, Kotsakis T, Piras P. 2016. *Condylura* (Mammalia, Talpidae) reloaded: New insights about the fossil representatives of the genus. *Palaeontol Electron.* 19.3.54A

1208

1209 Sansalone G, Colangelo P, Loy A, Raia P, Wroe S, Piras P. 2019. Impact of transition to a
1210 subterranean lifestyle on morphological disparity and integration in talpid moles (Mammalia,
1211 Talpidae). *BMC Evol Biol.* 19:179.

1212

1213 Scholander P. 1962. Physiological adaptation to diving in animals and man. *Harvey Lect.* 57:93–
1214 110.

1215

1216 Schwermann AH, Thompson RS. 2015. Extraordinarily preserved talpids (Mammalia,
1217 Lipotyphla) and the evolution of fossoriality. *J Vert Paleontol.* 35:e934828.

1218

1219 Simmons MP, Kessenich J. 2020. Divergence and support among slightly suboptimal likelihood
1220 gene trees. *Cladistics* 36:322–340.

1221

1222 Simmons MP, Sloan DB, Springer MS, Gatesy J. 2019. Gene-wise resampling outperforms site-
1223 wise resampling in phylogenetic coalescence analyses. *Mol Phylogenet Evol.* 131:80–92.

1224

1225 Simpson JT, Wong K, Jackman SD, Schein JE, Jones SJM, Birol I. 2009. ABySS: A parallel
1226 assembler for short read sequence data. *Genome Res.* 19:1117–1123.

1227 Springer MS, Murphy WJ, Roca AL. 2018. Appropriate fossil calibrations and tree constraints
1228 uphold the Mesozoic divergence of solenodons from other extant mammals. *Mol Phylogent Evol.*
1229 121:158–165.

1230

1231 Springer MS, Stanhope MJ, Madsen O, de Jong WW. 2004. Molecules consolidate the placental
1232 mammal tree. *Trend Ecol Evol.* 19:430–438.

1233

1234 Stamatakis A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of
1235 large phylogenies. *Bioinformatics* 30:1312–1313.

1236

1237 Stöver BC, Müller KF. 2010. TreeGraph 2: Combining and visualizing evidence from different
1238 phylogenetic analyses. *BMC Bioinformatics* 11:7.

Vogel P, Bodmer C, Spreng M, Aeschimann J. 2007. Diving capacity and foraging behaviour of the water shrew. In: Gorman ML, Dunstone N, editors. Behaviour and Ecology of Riparian Mammals. New York: Cambridge University Press Cambridge, United Kingdom. p. 31–48.

Waterhouse A, Bertoni M, Bienert S, Studer G, Tauriello G, Gumienny R, Heer FT, de Beer TAP, Rempfer C, Bordoli L, Lepore R, Schwede T. 2018. SWISS-MODEL: homology modelling of protein structures and complexes. *Nucleic Acids Res.* 46:W296–W303.

Whidden HP. 1999. The evolution of locomotor specializations in moles. *Am Zool.* 39: 135A.

Whidden HP. 2000. Comparative myology of moles and the phylogeny of the Talpidae (Mammalia, Lipotyphla). *Am Mus Novit.* 3294:1–53.

Woodman N, Gaffney SA. 2014. Can they dig it? Functional morphology and semifossoriality among small-eared shrews, genus *Cryptotis* (Mammalia, Soricidae). *J Morphol.* 275:745–759.

Yang Z. 2007. PAML 4: phylogenetic analysis by maximum likelihood. *Mol Biol Evol.* 24:1586–1591.

Zhang C, Rabiee M, Sayyari E, Mirarab S. 2018. ASTRAL-III: polynomial time species tree reconstruction from partially resolved gene trees. *BMC Bioinformatics* 19:153.

Zimin AV, Smith DR, Sutton G, Yorke JA. 2008. Assembly reconciliation. *Bioinformatics* 24:42–45.