MACROEVOLUTION WITH FOSSIL AND LIVING TAXA

by

THOMAS GUILLERME

B.Sc., Université Montpellier 2, 2010
M.Sc., Université Montpellier 2, 2012

A thesis submitted in partial fulfillment of
the requirements for the degree of

DOCTOR OF PHILOSOPHY

School of Natural Sciences
(Zoology)

Trinity College Dublin

SEPTEMBER 2015

DECLARATION

I declare that this thesis has not been submitted as an exercise for a degree at this or any other University and it is, unless otherwise referenced, entirely my own work. I agree to deposit this thesis in the University's open access institutional repository or allow the library to do so on my behalf, subject to Irish Copyright Legislation and Trinity College Library conditions of use and acknowledgement.

Thomas Guillerme

Summary

Even if most of our current knowledge and tool-kits to study biodiversity focus on living species, the vast majority of the species that ever lived are long extinct. Therefore, to properly understand the drivers of biodiversity through time, it is crucial to combine data and methods from both living and fossil species in order to better assess macroevolutionary and macroecological patterns. My PhD focuses on ways to combine both living and fossil species into phylogenies and looks at how these phylogenies can be used for describing macroevolutionary patterns. I studied the use of both living and fossil species along two axes: firstly, the ability of modern phylogenetic methods to deal with molecular data for living species and morphological data for both living and fossil species; and secondly, the practicality of using such phylogenetic trees for more accurately describing patterns of diversification through time and space.
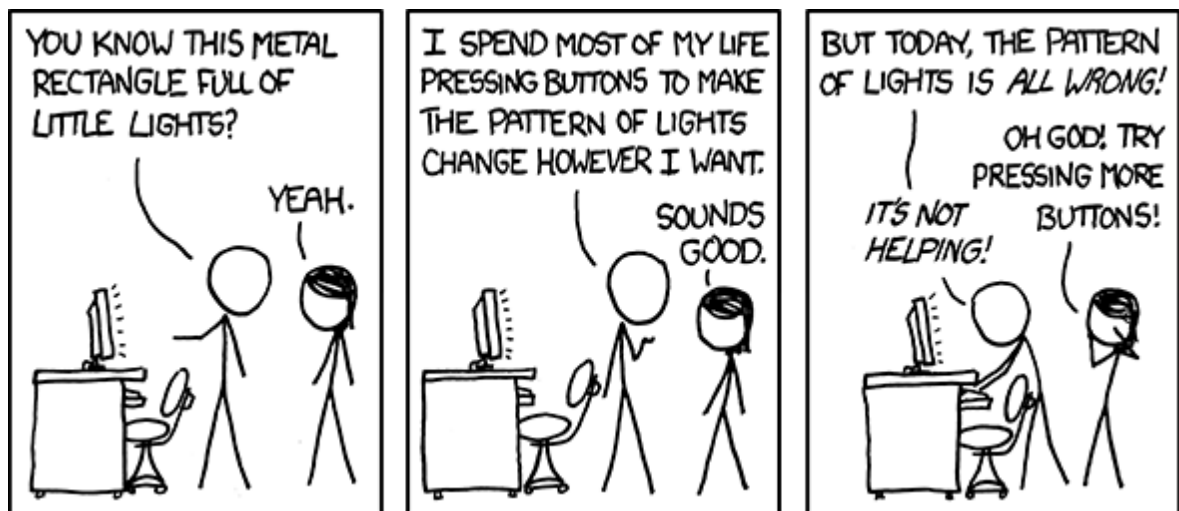
For the first part of this project, I ran extensive and thorough simulation analyses to test the effect of missing data on phylogenies topologies when using a combination of living and fossil data. I tested how multiple levels of missing data among living species, fossil species and the two combined affected our ability to recover the correct tree topology. I found that the amount of missing data among living species is the most crucial aspect for efficiently combining living and fossil species in the same phylogeny. Following these conclusions, I ran a thorough survey of the data available for living mammal species. I measured the amount of morphological data available within each mammalian order and tested whether this data was randomly distributed along the phylogeny or biased towards certain clades. The result of this analysis shows that although morphological data is scarce for living mammals, it is at least generally randomly distributed across the phylogeny.

For the second part of my PhD, I explored a way of using phylogenetic trees containing both living and fossil species to measure patterns of diversification among mammals through time. I measured changes in species richness as well as in morphological diversity (i.e. disparity) to describe mammalian diversification across the K-Pg boundary. I found that the K-Pg boundary had no significant effect on morphological diversification.

ACKNOWLEDGEMENTS

Thanks folks!

xkcd.com/722 - CC BY-NC 2.5

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Specific chapter contributions and acknowledgements

Effects of missing data on topological inference using a total evidence approach

Data availability and reproducibility

All the code used in this analysis is available on GitHub (goo.gl/4djNUf) with some information on how to use the various functions. Additionally all the simulated data is available on FigShare (dx.doi.org/10.6084/m9.figshare.1306861).

Contributions

I designed the experiment, ran the analysis, interpreted the results and wrote the manuscript. Natalie Cooper helped for designing the experiment and writing the manuscript.

Missing data in living mammals

Data availability and reproducibility

All data and analysis code is available on GitHub (`https://github.com/TGuillerme/Missing_living_mammals`).

CRETACEOUS-PALAEOGENE EXTINCTION DOES NOT AFFECT MAMMALIAN DISPARITY

Data availability and reproducibility

Data will be available on Dryad or Figshare. Code for reproducing the analysis is available on GitHub (`ithub.com/TGuillerme/SpatioTemporal_Disparity`).

INTRODUCTION

The amazing diversity of organisms living on the biosphere represents an over-whelmingly low fraction of the organisms that ever existed (**??**), yet, most of the work in biology focus solely on living species (Fritz et al., 2013). Ignoring this, can lead to misinterpretation of macroevolutionary or macroecological patterns (**???**Slater and Harmon, 2013; Fritz et al., 2013; Benton, 2015).

Although most species that have ever lived are now extinct (**??**), the majority of macroevolutionary studies focus solely on living species (e.g. Meredith et al., 2011; **?**)

Studying events in the evolutionary history of a taxonomic group, such as adaptive radiation or extinction, requires a fine-Â■âĂŘscale and accurate resolution of their phylogenetic relationships through time. To achieve this, most scientists would agree that information about both extant and extinct species is needed. However, few efforts have been made to combine extant and extinct species in the same phylogenetic trees; instead phylogenetic trees usually contain only extant species. Because the vast majority of species in a lineage will be represented by extinct species, studies focusing on extant species contain less than 0.1% of the lineageâĂŹs species richness. In some clades, ignoring extinct species may also obscure the true evolutionary history, species richness (i.e. Proboscidea), biogeography (i.e. Tinamiforms) or ecological diversity (i.e. Crocodilomorphs). Thus including the fossil record in these kinds of studies is essential to fully understand the evolutionary history of lineages.

In some clades, ignoring extinct species may also obscure the true evolutionary history or diversity of a clade. For example, in mammals the orders Perissodactyla (horses, rhinos and tapirs) and Proboscidea (elephants and mammoths) currently contain only a few taxa although they contained many more species in the past (Cifelli 1981, Antoine et al. 2003, Mihlbachler 2008, Seiffert et al. 2012). Crocodilo-morph reptiles also had much greater species diversity (there were 45 genera during the Paleogene1 but only nine today - Martin 2009) and larger geographic ranges (e.g. extinct Crocodilomorpha remains are found in northern France - Hua 1997) in the

past. Finally, ignoring extinct diversity can lead to misinterpretation in comparative phylogenetic studies. For example, ignoring extinct Dinornithiforms (Moas) leads to misinterpretation of both the diversification pattern and biogeography of ratites (Jetz et al. 2012). Thus including the fossil record in these kinds of studies is essential to fully understand the evolutionary history of lineages (Slater et al. 2012).

1.1 COMBINING DATA, METHODS AND DISCIPLINES

grab and expand the examples from the introductions.

The traditional way to combine both living and fossils data is using morphological data for both and draw conclusions on want happened. However the problem is that such approach simply exclude living species (e.g. trilobites) or use living species just as a way to branch the fossil species in a macroevolutionary context (e.g. O'Leary or any other big cladistic paper?). Also, the methods use to describe relations can be have big artefacts (e.g. parsimony) or other approaches (Mk) model can be oversimplistic (but still usuable). Finally the data for fossils is usually restricted to morphology and the few ecological traits that can be extract from that (e.g. diet from teeth but not behaviour or population size) Another approach for looking at macroevolution is to solely use living species (e.g. Jetz) one can look at the differences between DNA (many differences, good) and use models that are more realistic than Mk because of only four states(e.g. GTR). Also this method can include time by calibrating the molecular clocks using fossils (Zuckerkandl). However, appart from the use of the fossils occurence dates for making the clock tick, this approach ignores all clades that have no living descendant (the majority of clades!) and can even poorlierly estimate things from fossils.

Therefore, combining both data allows use to palliate to some of the problems of both approaches!

1.2 PHYLOGENIES WITH LIVING AND FOSSIL SPECIES

TEM Intro on the data story.

1.2.1 *Effects of missing data on topological inference using a Total Evidence approach*

In the first chapter, I run long term and thorough simulation to test how robust are our phylogenetic inferences when we combine living taxa with molecular and morphological to fossil taxa with morphological data only. I particularly focused on how missing data in both living and fossil taxa can affect topology. I found that the number of living taxa with available data is essential to recover accurate topologies. Therefore we need data for living mammals, how much of it is out there? (This chapter is currently in review in Molecular Phylogenetics and Evolution - revisions).

1.2.2 *Morphological data availability in living mammals*

Following these results, I was interested in showing practical implications of this effect and monitored the morphological data availability for living mammals. I downloaded all the recent available morphological matrices and counted the number of living mammals with available morphological data. I then tested how these taxa where distributed accross the phylogeny to check if there weren't clustered in some specific clades. I found that a lot of data is missing but that at least most of it is randomly distributed and should not drastically effect topology. So it's not so bad, but what can we do with these total evidence trees? (This chapter is an invited submission to a special issue in Biology Letters. Submission is due in December 2015).

1.3 TOTAL EVIDENCE PHYLOGENIES APPLICATIONS

Once we have these trees we can do loads of cool stuff (primates ideas).

1.3.1 *Cretaceous-Palaeogene extinction does not affect mammalian disparity*

One important step in explaining macroevolutionnary processes is to accurately describe the patterns. For example, to explain the processes driving diversification or extinction during a mass extinction event, we need to accurately measure what's happening. One classical example is the K-T extinction where the effects still remain unclear after so many years of research. Because I showed in the previous chapters that mammals are ok for combinations, I studied how they get affected by K-T

disparity wise (INTRODUCE DISPARITY FIRST). I found that... All the cool stuff we can do with TEM! (This chapter will be submitted to Evolution).

1.4 FURTHER DIRECTIONS

Finally I will discuss all these cool stuff and how research might develop by doing the combinations of data and looking at more accurate descriptors of patterns (e.g. diversity AND disparity) blalbalblabla.

Effects of missing data on topological inference using a Total Evidence approach[1]

---

[1]A similar version of this chapter is currently (2015/09/30) under review in Molecular Phylogenetics and Evolution.

Assessment of cladistic data availability for living mammals[2]

3.1 SUMMARY

Analyses of living and fossil taxa are crucial for understanding changes in biodiversity through time. The Total Evidence method allows living and fossil taxa to be combined in phylogenies, by using molecular data for living taxa and morphological data for both living and fossil taxa. With this method, substantial overlap of morphological data among living and fossil taxa is crucial for accurately inferring topology. However, although molecular data for living species is widely available, scientists using and generating morphological data mainly focus on fossils. Therefore, there is a gap in our knowledge of neontological morphological data even in well-studied groups such as mammals.

We investigated the amount of morphological (cladistic) data available for living mammals and how this data was phylogenetically distributed across orders. 22 of 28 mammalian orders have <25% species with available morphological data; this has implications for the accurate placement of fossil taxa, although the issue is less pronounced at higher taxonomic levels. In most orders, species with available data are randomly distributed across the phylogeny, which may reduce the impact of the problem. We suggest that increased morphological data collection efforts for living taxa are needed to produce accurate Total Evidence phylogenies.

---

[2]A shorter version (2500 words) will be submitted to Biology Letters as an invited submission for a special issue on phylogenies with living and fossil species. This special issue is open to submission in December 2015.

3.2 INTRODUCTION

There is an increasing consensus among evolutionary biologists that studying both living and fossil taxa is essential for fully understanding macroevolutionary patterns and processes (Slater and Harmon, 2013; Fritz et al., 2013; Wood et al., 2013). For example, including both living and fossil taxa in evolutionary studies can improve the accuracy of timing diversification events (e.g. Ronquist et al., 2012), our understanding of relationships among lineages (e.g. Beck and Lee, 2014), and our ability to infer biogeographical patterns through time (e.g. Meseguer et al., 2015). To perform such analyses it is necessary to combine living and fossil taxa in phylogenetic trees. One increasingly popular method, the Total Evidence method (Eernisse and Kluge, 1993; Ronquist et al., 2012), combines molecular data from living taxa and morphological data from both living and fossil taxa in a supermatrix (e.g. Pyron, 2011; Ronquist et al., 2012; Schrago et al., 2013; Slater and Harmon, 2013; Beck and Lee, 2014; Meseguer et al., 2015), producing a phylogeny with living and fossil taxa at the tips. These phylogenies can be dated using methods such as tip-dating (Ronquist et al., 2012; Wood et al., 2013) and incorporated into macroevolutionary studies (e.g. Ronquist et al., 2012; Wood et al., 2013; Slater, 2013).

A downside of the Total Evidence method is that it requires a lot of data. One must collect molecular data for living taxa and morphological data for both living and fossil taxa; two types of data that require fairly different technical skills (e.g. molecular sequencing *vs.* anatomical description). Additionally, large chunks of this data can be difficult, or even impossible, to collect for every taxon present in the analysis. For example, fossils very rarely have molecular data and incomplete fossil preservation (e.g. soft *vs.* hard tissues) may restrict the amount of morphological data available (Sansom and Wills, 2013). Additionally, since the molecular phylogenetics revolution, it has become less common to collect morphological characters for living taxa when molecular data is available (e.g. in (Slater, 2013), only 13% of the 169 living taxa have coded morphological data). Unfortunately this missing data can lead to errors in phylogenetic inference; in fact, simulations show that the ability of the Total Evidence method to recover the correct phylogenetic topology decreases when there is a low overlap between morphological data in the living and fossil taxa (Guillerme and Cooper, 2015), regardless the overall amount of morphological data available for the fossils (or the amount of molecular data available for the living species). The effect of missing data on topology is greatest when living taxa have

few morphological data. This is because (1) a fossil cannot branch in the correct clade if there is no overlapping morphological data in the clade; and (2) a fossil has a higher probability of branching within a clade with more morphological data available for living taxa, regardless of whether this is the correct clade or not (Guillerme and Cooper, 2015).

The issues above highlight that it is crucial to have sufficient morphological data for living taxa in a clade before using a Total Evidence approach. However, it is unclear how much morphological data for living taxa is actually available (i.e. already coded from museum specimens and deposited in phylogenetic matrices accessible online), and how this data is distributed across clades. Intuitively, most people assume this kind of data has already been collected, but empirical data suggest otherwise (e.g. in (Ronquist et al., 2012; Slater, 2013; Beck and Lee, 2014). To investigate this further, we assess the amount of available morphological data for living mammals to determine whether sufficient data exists to build reliable Total Evidence phylogenies in this group. We collected cladistic data (i.e. discrete morphological characters used in phylogenetics) from 286 phylogenetic matrices available online and measured the proportion of cladistic data available for each mammalian order. Additionally, because missing morphological data in living species can influence tree topology as described above, we determined whether the available cladistic data was phylogenetically overdispersed or clustered in the mammalian orders where data was missing. We find that available morphological data for living mammals is scarce but generally randomly distributed across phylogenies. We recommend that efforts be made to collect and share more cladistic data for living species to improve the accuracy of Total Evidence phylogenies.

3.3 MATERIALS AND METHODS

3.3.1 *Data collection and standardisation*

We downloaded all cladistic matrices containing any living and/or fossil mammal taxa from three major public databases (accessed 10th of June 2015): Morphobank (`http://www.morphobank.org/` (O'Leary and Kaufman, 2011), Graeme Lloyd's website (`graemetlloyd.com/matrmamm.html` and Ross Mounce's GitHub repository (`https://github.com/rossmounce/cladistic-data`. We also performed a systematic Google Scholar search (accessed 11th of June 2015) for matrices that were not uploaded to

these databases. We downloaded available matrices containing fossil and/or living mammal taxa from the three data bases using the following list of keywords:

Mammalia; Monotremata; Marsupialia; Placentalia; Macroscelidea; Afrosoricida; Tubulidentata; Hyracoidea; Proboscidea; Sirenia; Pilosa; Cingulata; Scandentia; Dermoptera; Primates; Lagomorpha; Rodentia; Erinaceomorpha; Soricomorpha; Cetacea; Artiodactyla; Cetartiodactyla; Chiroptera; Perissodactyla; Pholidota; Carnivora; Didelphimorphia; Paucituberculata; Microbiotheria; Dasyuromorphia; Peramelemorphia; Notoryctemorphia; Diprotodontia.

Note that some matrices have been downloaded from more than one database but that it is not an issue since we are interested in the total number of unique living OTUs and that if some where present in more than one matrix, they still only counted as one single OTU.

MORPHOBANK — We used the keywords listed above in the search menu of the Morphobank repository (`http://www.morphobank.org/`) and downloaded the data associated with each project matching with the keywords.

GRAEME LLOYD — We downloaded all the matrices that were available with a direct download link in the mammal data section of Graeme Lloyd's website repository (`http://graemetlloyd.com/`).

ROSS MOUNCE — We downloaded every 601 matrix from Ross Mounce's GitHub repository (`https://github.com/rossmounce`) and then ran a shell script to select only the matrices that had any text element that match with one of the search terms. To make the matrix selection more thorough, we ignored the keywords case as well as the latin suffix (*ia*, *ata*, *ea*, and *a*).

GOOGLE SCHOLARS (ACCESSED 11TH OF JUNE 2015) — To make sure we didn't miss any extra matrix that wasn't available on one of these repository, we ran a extra Google Scholar search. We downloaded the additional cladistic matrices from the 20 first search results matching with our selected keywords and with any of the 35 taxonomic levels (mammals Orders, Infraclasses and Class). We used the following key words:

order ("morphology" OR "morphological" OR "cladistic") AND characters matrix paleontology phylogeny

were *order* was replaced by all the keywords listed above. For each 33 keywords, we selected the 20 first papers to match the Google search published since 2010 resulting in 660 papers. Among these papers, not all contained relevant data (discrete morphological characters AND mammalian data). We selected only the 20 first results per search term to avoid downloading articles that were to irrelevant. Among the 660 papers, only 50 contained a total of 425 extra living OTUs (Figure 3.1). Also we decided to select only the articles published since 2010 because nearly every one of the recent published matrix contains both a fraction of morphological characters and OTUs from previous studies. For example in primates the character *p7* coded first by Ross et al. (1998) is reused with the same living species in Seiffert et al. (2003), Marivaux et al. (2005), Seiffert et al. (2005), Bloch et al. (2007), Bloch et al. (2007), Kay et al. (2008), Silcox (2008), Seiffert et al. (2009), Tabuce et al. (2009), Boyer et al. (2010), Seiffert et al. (2010), Marivaux et al. (2013) and Ni et al. (2013).

We transformed all the non-nexus matrices (tnt, word, excel, jpeg) to nexus format manually. In total, we downloaded 286 matrices containing a total of 11010 operational taxonomic units (OTUs) of which 5228 were unique. In this study, we refer to OTUs rather than species since the entries in the downloaded matrices were not standardised and ranged from specific individual specimen names (i.e. the name of a collection item) to the family-level. Where possible, we considered OTUs at their lowest valid taxonomic level (i.e. species) but some OTUs were only valid at a higher taxonomic level (e.g. genus or family). Therefore for some orders, we sampled more genera than species (Table **??**.

To select the lowest valid taxonomic level for each OTU, we standardised their taxonomy by correcting species names so they matched standard taxonomic nomenclature (e.g., *H. sapiens* was transformed to *Homo sapiens*. We designated as "living" all OTUs that were either present in the phylogeny of (Bininda-Emonds et al., 2007) or the taxonomy of (Wilson and Reeder, 2005), and designated as "fossil" all OTUs that were present in the Paleobiology database (`https://paleobiodb.org/`.

For OTUs that did not appear in these three sources, we first decomposed the name (i.e. *Homo sapiens* became *Homo* and *sapiens* and tried to match the first element with a higher taxonomic level (family, genus etc.). Any OTUs that still had no matches in the sources above were designated as non-applicable (NA; see Figure 3.2).
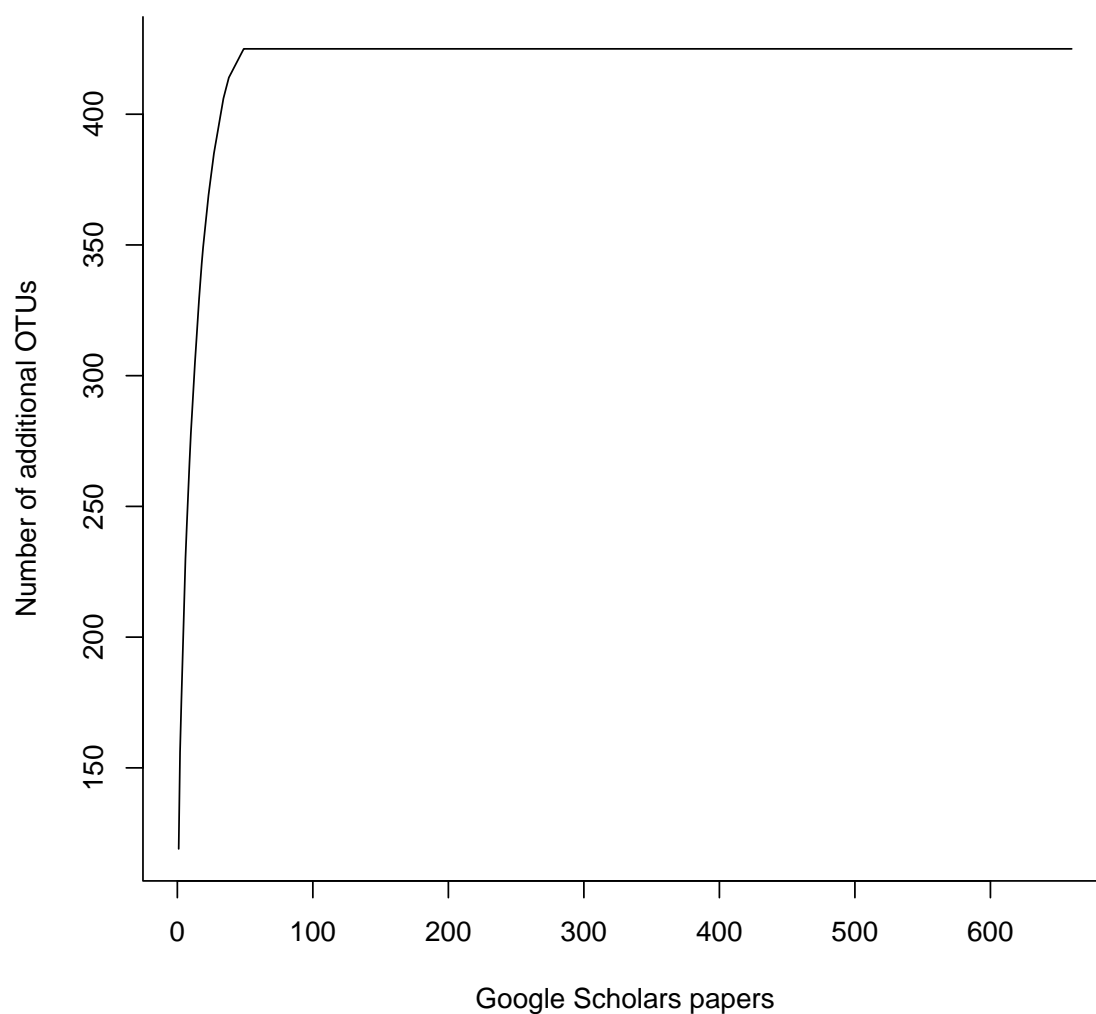
FIGURE 3.1: Google searches additional OTUs rarefaction curve. The x axis represent the number of google scholar matches (papers, books or abstracts) and the y axis represents the cumulative number of additional living OTUs per google scholar match.
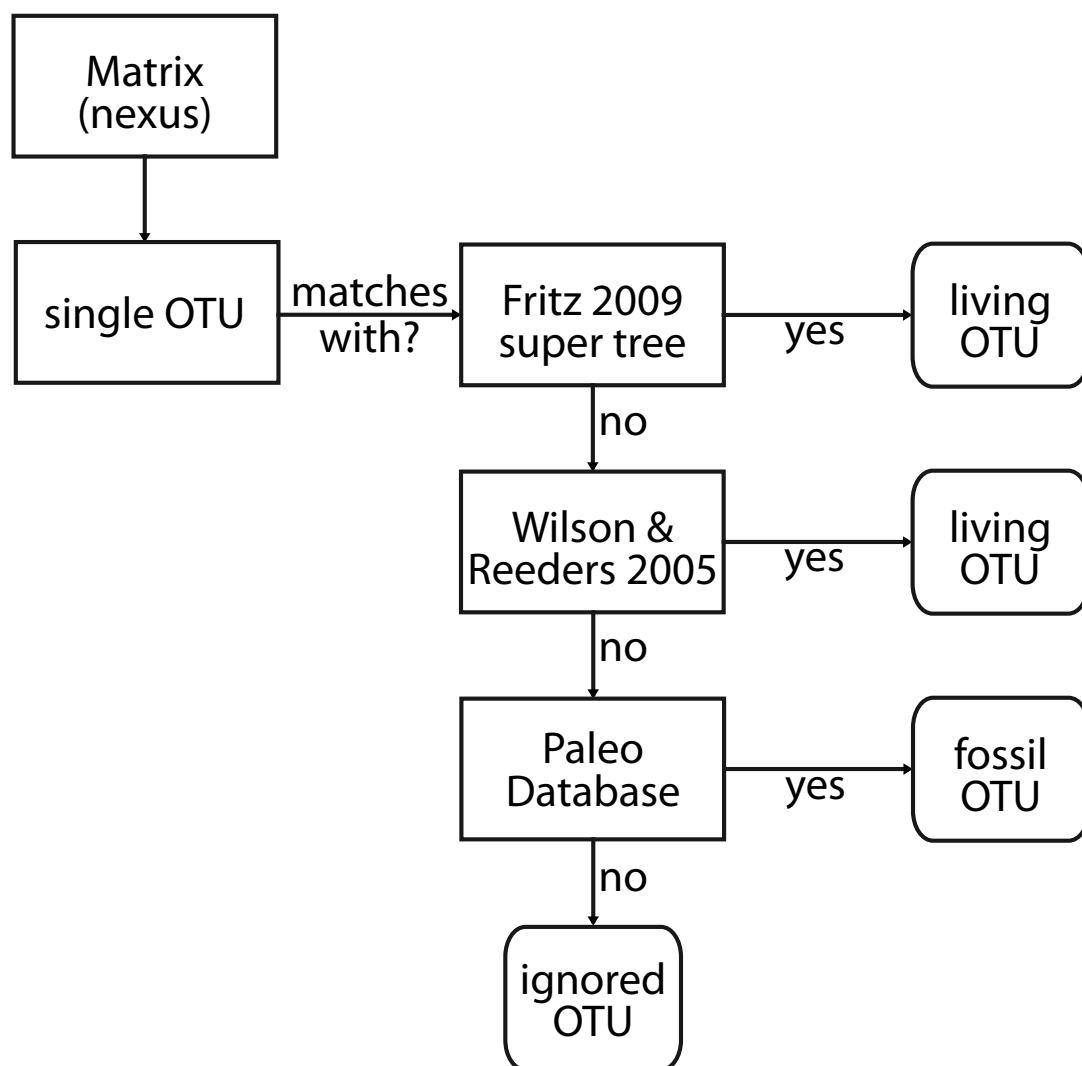
FIGURE 3.2: Taxonomic matching algorithm used in this study. For each matrix, each operational taxonomic units (OTU) is matched with the super tree from Bininda-Emonds 2007. If the OTU matches, then it is classified as living. Else it is matched with the Wilson & Reeders 2005 taxonomy list. If the OTU matches, then it is classified as living. Else it is matched with the Paleo Database list of mammals. If the OTU matches, then it is classified as fossil. Else it is ignored.

The number of characters in each matrix ranged from 6 to 4541. Matrices with few characters are problematic when comparing available data among matrices because (1) they have less chance of having characters that overlap with those of other matrices (Wagner, 2000) and (2) they are more likely to contain a higher proportion of specific characters that are not-applicable across large clades (e.g. "antler ramifications" is a character that is only applicable to Cervidae not all mammals (Brazeau, 2011). Therefore we selected only matrices containing >100 characters for each OTU. This threshold was chosen to correspond with the number of characters used in (Guillerme and Cooper, 2015) and (Harrison and Larsson, 2015). Note that results of analyses with no character threshold are available in Supplementary Material. After removing all matrices with <100 characters, we retained 1074 unique living mammal OTUs from 126 matrices for our analyses.

3.3.2 *Data availability and distribution*

To assess the availability of cladistic data for each mammalian order, we calculated the percentage of OTUs with cladistic data at three different taxonomic levels: family, genus and species. We consider orders with <25% of living taxa with cladistic data as having poor data coverage ("low" coverage), and orders with >75% of living taxa with cladistic data as having good data coverage (hereafter "high" coverage).

For orders with <100% cladistic data coverage at any taxonomic level, we investigated whether the available cladistic data was (i) randomly distributed, (ii) overdispersed or (iii) clustered, with respect to phylogeny, using two metrics from community phylogenetics: the Nearest Taxon Index (NTI; (Webb et al., 2002) and the Net Relatedness Index (NRI; (Webb et al., 2002). NTI is most sensitive to clustering or overdispersion near the tips, whereas NRI is more sensitive to clustering or overdispersion across the whole phylogeny (Cooper et al., 2008). Both metrics were calculated using the `picante` package in R (Kembel et al., 2010; R Core Team, 2015).

NTI (Webb et al., 2002) is based on mean nearest neighbour distance (MNND) and is calculated as follows:

$$NTI = - \left(\frac{\overline{MNND}_{obs} - \overline{MNND}_n}{\sigma(MNND_n)} \right) \tag{3.1}$$

where $\overline{MNND}_{obs}$ is the observed mean distance between each of n taxa with cladistic data and its nearest neighbour with cladistic data in the phylogeny, $\overline{MNND}_n$ is

the mean of 1000 mean MNND between n randomly drawn taxa, and $\sigma(MNND_n$ is the standard deviation of these 1000 random MNND values. NRI is similar but is based on mean phylogenetic distance (MPD) as follows:

$$NRI = -\left(\frac{\overline{MPD}_{obs} - \overline{MPD}_n}{\sigma(MPD_n}\right) \tag{3.2}$$

where $\overline{MPD}_{obs}$ is the observed mean phylogenetic distance of the tree containing only the n taxa with cladistic data, $\overline{MPD}_n$ is the expected random MPD for n taxa estimated by calculating the MPD from n taxa randomly drawn from the phylogeny and repeated 1000 times, and $\sigma(MPD_n$ is the standard deviation of the 1000 random MPD values. Negative NTI and NRI values show that the focal taxa are more overdispersed across the phylogeny than expected by chance, and positive values reflect significant clustering.

We calculated NTI and NRI values for each mammalian order separately, at each different taxonomic level. For each analysis our focal taxa were those with available cladistic data at that taxonomic level and the phylogeny was the phylogeny of the order pruned from (Bininda-Emonds et al., 2007).
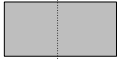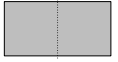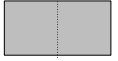
3.4 RESULTS
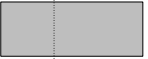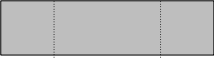
Across the 126 cladistic matrices we extracted, 22 out of 28 mammalian orders have low coverage (<25% of species with cladistic data) and six have high coverage (>75% of species with cladistic data) at the species-level. At the genus-level, three orders have low coverage and 12 have high coverage, and at the family-level, no orders have low coverage and 23 have high coverage (Table **??**.
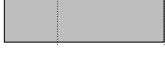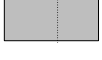
TABLE 3.1: Number of taxa with available cladistic data for mammalian orders at three taxonomic levels. The left vertical bar represents "low" coverage (<25%); the right vertical bar represents "high" coverage (>75%). A negative Net Relatedness Index (NRI) and Nearest Taxon Index (NTI) shows more phylogeneticaly dispersed taxa than expected by chance; a positive value shows more phylogeneticaly clustered taxa than expected by chance. Significant NRI or NTI values are highlighted in bold. One star (*) signifies a p-value between 0.05 and 0.005; two starts between 0.005 and 0.0005 and three stars <0.0005.

Order	Taxonomic level	Proportion of taxa	Coverage	NRI	NTI

Afrosoricida	family	2/2			
Afrosoricida	genus	17/17			
Afrosoricida	**species**	**23/42**		**1.89***	1.19
Carnivora	family	11/15		0.43	1.68
Carnivora	**genus**	**30/125**		**4.14****	**1.81***
Carnivora	**species**	**42/283**		**18.64****	**3.02****
Cetartiodactyla	family	21/21			
Cetartiodactyla	**genus**	**77/128**		0.87	**1.77***
Cetartiodactyla	**species**	**129/310**		**2.72***	0.04
Chiroptera	family	13/18		0.55	0.63
Chiroptera	**genus**	**85/202**		**16.91****	**2.85****
Chiroptera	**species**	**165/1053**		**14.55****	**3.44****
Cingulata	family	1/1			
Cingulata	genus	8/9		1.49	-1.63
Cingulata	species	6/29		1.43	0.36
Dasyuromorphia	family	2/2			
Dasyuromorphia	genus	7/22		-1	-1.45
Dasyuromorphia	species	8/64		-1.15	-0.62
Dermoptera	family	1/1			
Dermoptera	genus	1/2			
Dermoptera	species	1/2			
Didelphimorphia	family	1/1			
Didelphimorphia	genus	16/16			
Didelphimorphia	species	40/84		-0.94	0.36

Diprotodontia	family	9/11		-0.8	0.56
Diprotodontia	genus	20/38		-1.36	-0.73
Diprotodontia	species	16/126		-2.29	-1.55
Erinaceomorpha	family	1/1			
Erinaceomorpha	genus	10/10			
Erinaceomorpha	species	21/22		-1.1	-0.3
Hyracoidea	family	1/1			
Hyracoidea	genus	1/3			
Hyracoidea	species	1/4			
Lagomorpha	family	1/2			
Lagomorpha	genus	1/12			
Lagomorpha	species	1/86			
Macroscelidea	family	1/1			
Macroscelidea	genus	4/4			
Macroscelidea	species	5/15		-0.98	-1.38
Microbiotheria	family	1/1			
Microbiotheria	genus	1/1			
Microbiotheria	species	1/1			
Monotremata	family	2/2			
Monotremata	genus	2/3		-0.71	-0.71
Monotremata	species	2/4		-1.01	-1.03
Notoryctemorphia	family	1/1			
Notoryctemorphia	genus	1/1			
Notoryctemorphia	species	0/2			

Taxon	Rank	Ratio		Value 1	Value 2
Paucituberculata	family	1/1			
Paucituberculata	genus	2/3		0	0
Paucituberculata	species	2/5		-0.64	-0.65
Peramelemorphia	family	2/2			
Peramelemorphia	genus	7/7			
Peramelemorphia	species	16/18		-0.09	1
Perissodactyla	family	3/3			
Perissodactyla	genus	6/6			
Perissodactyla	species	7/16		0.62	-2.5
Pholidota	family	1/1			
Pholidota	genus	1/1			
Pholidota	**species**	**3/8**		**2.64***	**2.23***
Pilosa	family	3/5		0.94	0.93
Pilosa	genus	3/5		-0.36	-0.31
Pilosa	species	3/29		0.33	0.79
Primates	family	15/15			
Primates	genus	48/68		-0.41	-1.4
Primates	species	56/351		-1.6	-2.04
Proboscidea	family	1/1			
Proboscidea	genus	1/2			
Proboscidea	species	1/3			
Rodentia	family	11/32		-0.46	-1.91
Rodentia	genus	21/450		-2.11	0.3
Rodentia	species	15/2094		-1.65	-2.55

Scandentia	family	2/2			
Scandentia	genus	2/5		-0.77	-0.76
Scandentia	species	2/20		-1.79	-1.99
Sirenia	family	2/2			
Sirenia	genus	2/2			
Sirenia	species	4/4			
Soricomorpha	family	3/4		-0.93	-0.92
Soricomorpha	**genus**	**19/43**		**6.98****	**2.49***
Soricomorpha	**species**	**19/392**		**13.19****	**3.89****
Tubulidentata	family	1/1			
Tubulidentata	genus	1/1			
Tubulidentata	species	1/1			

Among the mammalian orders containing OTUs with no available cladistic data, only six orders had significantly clustered data (Carnivora, Cetartiodactyla, Chiroptera and Soricomorpha at both species- and genus-level and Afrosoricida and Pholidota at the species-level only) and no order had significantly overdispersed data at any taxonomic level (Table **??**.

Two contrasting results are shown in Figure 3.3 with randomly distributed OTUs with cladistic data in Primates (Figure 3.3A) and phylogenetically clustered OTUs with cladistic data in Carnivora (mainly Canidae; Figure 3.3B).

3.5   DISCUSSION

Our results show that although phylogenetic relationships among living mammals are well-resolved (e.g. Bininda-Emonds et al., 2007; Meredith et al., 2011) , most of the data used to build these phylogenies is molecular, and very little cladistic data is available for living mammals compared to fossil mammals (e.g. O'Leary et al., 2013; Ni et al., 2013). This has implications for building Total Evidence phylogenies
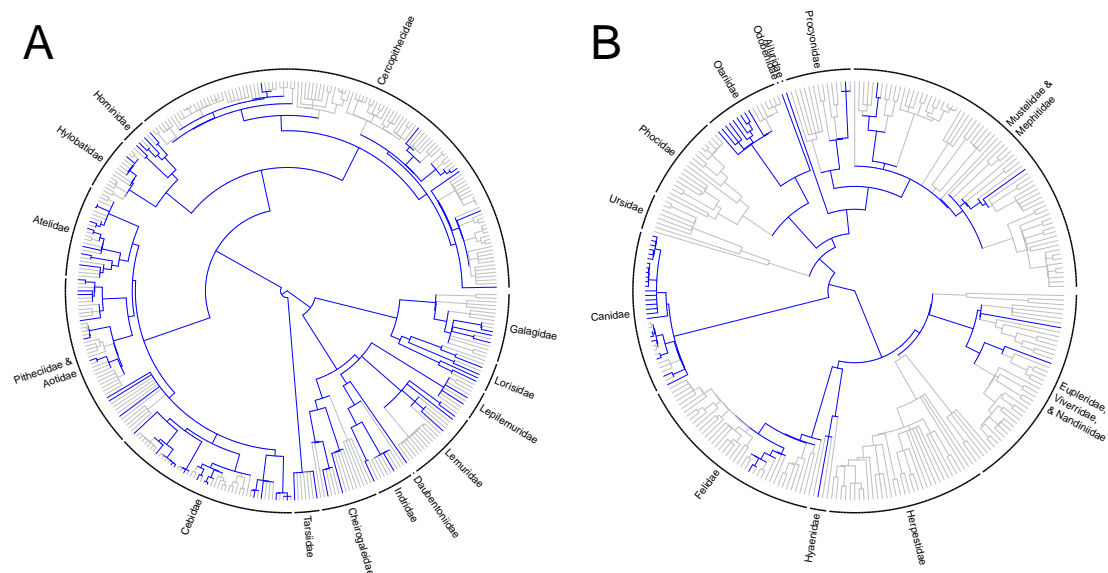
FIGURE 3.3: Phylogenetic distribution of species with available cladistic data across two mammalian orders (A: Primates; B: Carnivora). Edges are colored in grey when no cladistic data is available for a species and in blue when data is available.

containing both living and fossil mammals, as without sufficient cladistic data for living species, fossil placements in these trees are very uncertain (Guillerme and Cooper, 2015). Cladistic data coverage in living mammals varies across taxonomic levels and in its phylogenetic distribution. Higher taxonomic levels are always better sampled than lower ones and within these taxonomic levels, the available data is mostly randomly distributed across the phylogeny, apart from in six orders).

The number of living mammalian taxa with no available cladistic data was surprisingly high at the species-level: only six out of 28 orders have a high coverage of taxa with available cladistic data (and two of the 28 orders are monospecific!). This high coverage threshold of 75% of taxa with available cladistic data represents the minimum amount of data required before missing data has a significant effect on the topology of Total Evidence trees (Guillerme and Cooper, 2015). Beyond this threshold, there is considerable displacement of wildcard taxa (*sensu* (Kearney, 2002) and decreases in clade conservation (Guillerme and Cooper, 2015). Therefore we expect a high probability of topological artefacts for the placement of fossil taxa at the species-level in most mammalian orders. However, data coverage seems to be less of an issue at higher taxonomic levels (i.e. genus- and family-level). This point is important from a practical point of view because of the slight discrepancy between the neontological and palaeontological concept of species. While neontological species are described using morphology, genetic distance, spatial distribution and even behaviour, palaeontological species can be based only on morphological, spatial and temporal data (e.g. Ni et al., 2013). Because of this, most palaeontological studies are using the genus as their smallest OTU (e.g. Ni et al., 2013; O'Leary et al., 2013). Thus data availability at the genus-level in living mammals should be our primary concern when aiming to build phylogenies of living and fossil taxa.

When only a few species with cladistic data are available, the ideal scenario is for them to be phylogenetically overdispersed (i.e. that there is data for at least every sub-clade) to maximize the possibilities of a fossil branching from the right clade. The second best scenario is that species with cladistic data are randomly distributed across the phylogeny. In this scenario we expect no special bias in the placement of the fossil (Guillerme and Cooper, 2015), it is therefore encouraging that for most orders, species with cladistic data were randomly distributed across the phylogeny of each order. The worst case scenario for fossil placement is that species with cladistic data are phylogenetically clustered. In this situation we expect two major biases to occur: first, the fossil will not be able to branch within a clade

containing no data, and second, the fossil will have a higher probability, at random, of branching within the clade containing most of the available data. This means that fossils with uncertain phylogenetic affinities (*incertae sedis* will have a higher probability of branching within the most sampled clade just by chance. Our results suggest that this may be an issue, at the genus-level, in Carnivora, Cetartiodactyla, Chiroptera and Soricomorpha. For example, a Carnivora fossil will be unable to branch in the Herpestidae that has no species with cladistic data, and will also have more chance to branch, randomly, within the Canidae clade than any other clade in Carnivora (Figure 3.3B). Thus, in Total Evidence trees, placements of some carnivoran fossils should considered with caution. In this study, we treated all cladistic matrices as equal in a similar way to molecular matrices. For example, if matrix A contained 100 characters for taxa X and Y, and matrix B contained 50 different characters for taxa X and Z, we assumed that both matrices can be combined in a supermatrix containing 150 independent characters for taxon X, 100 for taxon Y and 50 for taxon Z. Unfortunately, cladistic data cannot always be treated in this way because some characters may overlap. For example, if matrix A has a character coding for the shape of a particular morphological feature and matrix B has a character coding for the presence of this same morphological feature and a second character coding for its shape, then these three characters are non-independent compound characters (Brazeau, 2011). However, in reasonably sized matrices (>100 characters; (Guillerme and Cooper, 2015; Harrison and Larsson, 2015) it is more likely that a number of characters are consistently conserved among the different matrices and thus easily combinable. For example, within the Primate cladistic literature, the character *p7* - the size of the 4th lower premolar paraconid - has been used consistently for >15 years (e.g. Ross et al., 1998; **?**; Ni et al., 2013) and can therefore be combined among the matrices. A conservative approach to avoid compound characters would be to select only the most recent matrix for each group, but this would result in the loss of a lot of data.

Despite the absence of good cladistic data coverage for living mammals, the Total Evidence methods still seems to be the most promising way of combining living and fossil data for macroevolutionary analyses. Following the recommendations in (Guillerme and Cooper, 2015), we need to code cladistic characters for as many living species possible. Fortunately, data for living mammals is usually readily available in natural history collections, therefore, we propose that an increased effort be put into coding morphological characters from living species, possibly by engaging

in collaborative data collection projects through web portals such as *Morphobank* (O'Leary and Kaufman, 2011). Such an effort would be valuable not only to phylogeneticists, but also to any researcher focusing understanding macroevolutionary patterns and processes.

Cretaceous-Palaeogene extinction does not affect mammalian disparity[3]

DISCUSSION

5.1 THE FUTURE OF THE TOTAL EVIDENCE METHOD

Combined with tip-dating is super interesting but: -Data limitations -Problems with dating (Arcila) -Better models for morphology? One way could be a REAL total evidence dating using also trait data, biogeography, etc... In reality, all this parameters have an influence of lineages history and should technically be taken into account. But data problem is likely to increase, an needs models need to be improved as well. And in the end, how many parameters do we want?

5.2 DIVERSITY IS MULTIDIMENSIONAL

It is important to disentangle But other dimensions as well: Ecological, life history, etc.

5.3 WHAT IS THE REAL EFFECT OF COMBINING?

Maybe only important when groups have actually a complex history? Old clades might have no living descendants and the question is therefore N/A Recent subclades maybe not have changed much in diversity so adding fossils might not change much. But we never know! Example of the giant lemur (recently extinct).

Future directions -what should we do on the thesis to continue

Caveats -the hard ones (I know this is broken) -phylogenies are never perfect -using proxies

Bibliography

Beck, R. M. and M. S. Lee. 2014. Ancient dates or accelerated rates? Morphological clocks and the antiquity of placental mammals. P. Roy. Soc. B-Biol. Sci. 281:1–10.

Benton, M. J. 2015. Exploring macroevolution using modern and fossil data. Proceedings of the Royal Society of London B: Biological Sciences 282.

Bininda-Emonds, O. R. P., M. Cardillo, K. E. Jones, R. D. E. MacPhee, R. M. D. Beck, R. Grenyer, S. A. Price, R. A. Vos, J. L. Gittleman, and A. Purvis. 2007. The delayed rise of present-day mammals. Nature 446:507–512.

Bloch, J. I., M. T. Silcox, D. M. Boyer, and E. J. Sargis. 2007. New paleocene skeletons and the relationship of plesiadapiforms to crown-clade primates. Proc. Nat. Acad. Sci. 104:1159–1164.

Boyer, D. M., E. R. Seiffert, and E. L. Simons. 2010. Astragalar morphology of afradapis, a large adapiform primate from the earliest late eocene of egypt. Am. J. Phys. Anthropol. 143:383–402.

Brazeau, M. D. 2011. Problematic character coding methods in morphology and their effects. Biol. J. Linn. Soc. 104:489–498.

Cooper, N., J. Rodríguez, and A. Purvis. 2008. A common tendency for phylogenetic overdispersion in mammalian assemblages. P. Roy. Soc. B-Biol. Sci. 275:2031–2037.

Eernisse, D. and A. Kluge. 1993. Taxonomic congruence versus total evidence, and amniote phylogeny inferred from fossils, molecules, and morphology. Mol. Biol. Evol. 10:1170–1195.

Fritz, S. A., J. Schnitzler, J. T. Eronen, C. Hof, K. Böhning-Gaese, and C. H. Graham. 2013. Diversity in time and space: wanted dead and alive. Trends Ecol. Evol. 28:509 – 516.

Guillerme, T. and N. Cooper. 2015. Effects of missing data on topological inference using a total evidence approach. Mol. Phylogenet. Evol. X:X.

Harrison, L. B. and H. C. E. Larsson. 2015. Among-character rate variation distributions in phylogenetic analysis of discrete morphological characters. Syst. Biol. 64:307–324.

Kay, R. F., J. Fleagle, T. Mitchell, M. Colbert, T. Bown, and D. W. Powers. 2008. The anatomy of dolichocebus gaimanensis, a stem platyrrhine monkey from argentina. J. Hum. Evol. 54:323–382.

Kearney, M. 2002. Fragmentary taxa, missing data, and ambiguity: mistaken assumptions and conclusions. Syst. Biol. 51:369–381.

Kembel, S., P. Cowan, M. Helmus, W. Cornwell, H. Morlon, D. Ackerly, S. Blomberg, and C. Webb. 2010. Picante: R tools for integrating phylogenies and ecology. Bioinformatics 26:1463–1464.

Marivaux, L., P.-O. Antoine, S. R. H. Baqri, M. Benammi, Y. Chaimanee, J.-Y. Crochet, D. De Franceschi, N. Iqbal, J.-J. Jaeger, G. Métais, et al. 2005. Anthropoid primates from the oligocene of pakistan (bugti hills): data on early anthropoid evolution and biogeography. Proc. Nat. Acad. Sci. 102:8436–8441.

Marivaux, L., A. Ramdarshan, E. M. Essid, W. Marzougui, H. K. Ammar, R. Lebrun, B. Marandat, G. Merzeraud, R. Tabuce, and M. Vianey-Liaud. 2013. Djebelemur, a tiny pre-tooth-combed primate from the eocene of tunisia: a glimpse into the origin of crown strepsirhines. PloS ONE 8:e80778.

Meredith, R., J. Janečka, J. Gatesy, O. Ryder, C. Fisher, E. Teeling, A. Goodbla, E. Eizirik, T. L. Simão, T. Stadler, D. Rabosky, R. Honeycutt, J. Flynn, C. Ingram, C. Steiner, T. Williams, T. Robinson, B. Angela, M. Westerman, N. Ayoub, M. Springer, and W. Murphy. 2011. Impacts of the Cretaceous terrestrial revolution and KPg extinction on mammal diversification. Science 334:521–524.

Meseguer, A. S., J. M. Lobo, R. Ree, D. J. Beerling, and I. Sanmartĩn. 2015. Integrating fossils, phylogenies, and niche models into biogeography to reveal ancient evolutionary history: The case of hypericum (hypericaceae). Systematic Biology 64:215–232.

Ni, X., D. L. Gebo, M. Dagosto, J. Meng, P. Tafforeau, J. J. Flynn, and K. C. Beard. 2013. The oldest known primate skeleton and early haplorhine evolution. Nature 498:60–64.

O'Leary, M. A., J. I. Bloch, J. J. Flynn, T. J. Gaudin, A. Giallombardo, N. P. Giannini, S. L. Goldberg, B. P. Kraatz, Z.-X. Luo, J. Meng, X. Ni, M. J. Novacek, F. A. Perini, Z. S. Randall, G. W. Rougier, E. J. Sargis, M. T. Silcox, N. B. Simmons, M. Spaulding, P. M. Velazco, M. Weksler, J. R. Wible, and A. L. Cirranello. 2013. The placental mammal ancestor and the postâĂŞK-Pg radiation of placentals. Science 339:662–667.

O'Leary, M. A. and S. Kaufman. 2011. Morphobank: phylophenomics in the cloud. Cladistics 27:529–537.

Pyron, R. 2011. Divergence time estimation using fossils as terminal taxa and the origins of Lissamphibia. Syst. Biol. 60:466–481.

R Core Team. 2015. R: a language and environment for statistical computing. R Foundation for Statistical Computing Vienna, Austria.

Ronquist, F., S. Klopfstein, L. Vilhelmsen, S. Schulmeister, D. Murray, and A. Rasnitsyn. 2012. A total-evidence approach to dating with fossils, applied to the early radiation of the Hymenoptera. Syst. Biol. 61:973–999.

Ross, C., B. Williams, and R. F. Kay. 1998. Phylogenetic analysis of anthropoid relationships. J. Hum. Evol. 35:221–306.

Sansom, R. S. and M. A. Wills. 2013. Fossilization causes organisms to appear erroneously primitive by distorting evolutionary trees. Scientific Reports 3:1–5.

Schrago, C., B. Mello, and A. Soares. 2013. Combining fossil and molecular data to date the diversification of New World Primates. J. Evolution. Biol. 26:2438–2446.

Seiffert, E. R., J. M. Perry, E. L. Simons, and D. M. Boyer. 2009. Convergent evolution of anthropoid-like adaptations in eocene adapiform primates. Nature 461:1118–1121.

Seiffert, E. R., E. L. Simons, and Y. Attia. 2003. Fossil evidence for an ancient divergence of lorises and galagos. Nature 422:421–424.

Seiffert, E. R., E. L. Simons, D. M. Boyer, J. M. Perry, T. M. Ryan, and H. M. Sallam. 2010. A fossil primate of uncertain affinities from the earliest late eocene of egypt. Proc. Nat. Acad. Sci. 107:9712–9717.

Seiffert, E. R., E. L. Simons, W. C. Clyde, J. B. Rossie, Y. Attia, T. M. Bown, P. Chatrath, and M. E. Mathison. 2005. Basal anthropoids from egypt and the antiquity of africa's higher primate radiation. Science 310:300–304.

Silcox, M. T. 2008. The biogeographic origins of primates and euprimates: east, west, north, or south of eden? Pages 199–231 *in* Mammalian Evolutionary Morphology. Springer.

Slater, G. J. 2013. Phylogenetic evidence for a shift in the mode of mammalian body size evolution at the cretaceous-palaeogene boundary. Methods Ecol. Evol. 4:734–744.

Slater, G. J. and L. J. Harmon. 2013. Unifying fossils and phylogenies for comparative analyses of diversification and trait evolution. Methods Ecol. Evol. 4:699–702.

Tabuce, R., L. Marivaux, R. Lebrun, M. Adaci, M. Bensalah, P.-H. Fabre, E. Fara, H. G. Rodrigues, L. Hautier, J.-J. Jaeger, et al. 2009. Anthropoid versus strepsirhine status of the african eocene primates algeripithecus and azibius: craniodental evidence. P. Roy. Soc. B-Biol. Sci.s Page rspb20091339.

Wagner, P. J. 2000. Exhaustion of morphologic character states among fossil taxa. Evolution 54:365–386.

Webb, C. O., D. D. Ackerly, M. A. McPeek, and M. J. Donoghue. 2002. Phylogenies and community ecology. Ann. Rev. Ecol. Syst. Pages 475–505.

Wilson, D. E. and D. M. Reeder. 2005. Mammal species of the world: a taxonomic and geographic reference vol. 1. JHU Press.

Wood, H. M., N. J. Matzke, R. G. Gillespie, and C. E. Griswold. 2013. Treating fossils as terminal taxa in divergence time estimation reveals ancient vicariance patterns in the palpimanoid spiders. Systematic Biology 62:264–284.

Assessment of cladistic data availability for living mammals
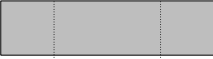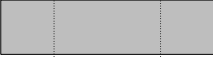
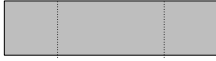Key words: Total Evidence method, data structure, phylogenetic, fossil, topology

A shorter version (2500 words) will be submitted to Biology Letters as an invited submission for a special issue on phylogenies with living and fossil species. This special issue is open to submission in December 2015.
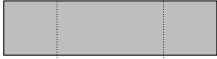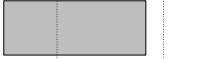
The following section contains supplementary results to the chapter "Assessment of cladistic data availability for living mammals": the available data structure using the NTI and the PD metric; the proportion of available data and the data structure for all the matrices (including the matrices with less than 100 characters); and phylogenetical representation of the data availability per order (excluding Primates and Carnivora, present in the main body).
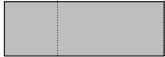
TABLE A.1: Number of taxa with available cladistic data for mammalian orders at three taxonomic levels (without any character threshold; results from the 286 matrices). The coverage represents the proportion of taxa with available morphological data. The left vertical bar represents 25% of available data ("low" coverage if <25%); The right vertical bar represents 75% of available data ("high" coverage if >75%). When the Net Relatedness Index (NRI) and the Nearest Taxon Index (NTI) are negative, taxa are more phylogenetically dispersed than expected by chance; when NRI or NTI are positive, taxa are more phylogenetically clustered by expected by chance. Significant NRI or NTI are highlighted in bold. One star (*) represents a p-value between 0.05 and 0.005; two starts between 0.005 and 0.0005 and three stars a p-value less than 0.0005.

Order	Taxonomic level	Proportion of taxa	Coverage		NRI	NTI
Afrosoricida	family	2/2				
Afrosoricida	genus	17/17				

Afrosoricida	species	23/42		1.75	1.08
Carnivora	family	14/15		0.63	0.6
Carnivora	**genus**	**54/125**		**4.81****	**1.78***
Carnivora	**species**	**76/283**		**7.66****	**0.85**
Cetartiodactyla	family	21/21			
Cetartiodactyla	genus	100/128		0.85	0.94
Cetartiodactyla	**species**	**171/310**		**1.92***	**-0.46**
Chiroptera	family	15/18		-0.28	0.56
Chiroptera	**genus**	**93/202**		**13.47****	**1.1**
Chiroptera	**species**	**215/1053**		**8.82****	**1.22**
Cingulata	family	1/1			
Cingulata	genus	8/9		1.51	-1.57
Cingulata	**species**	**9/29**		**1.9***	**0.11**
Dasyuromorphia	family	2/2			
Dasyuromorphia	genus	8/22		-0.75	-1.07
Dasyuromorphia	species	9/64		-0.88	-0.34
Dermoptera	family	1/1			
Dermoptera	genus	1/2			
Dermoptera	species	1/2			
Didelphimorphia	family	1/1			
Didelphimorphia	genus	16/16			
Didelphimorphia	species	42/84		-1.65	0.2
Diprotodontia	family	11/11			
Diprotodontia	genus	25/38		-1.13	-1.31

Taxon	Rank	Ratio	Bar	Value 1	Value 2
Diprotodontia	species	31/126		0.48	-1.77
Erinaceomorpha	family	1/1			
Erinaceomorpha	genus	10/10			
Erinaceomorpha	species	21/22		-1.07	-0.2
Hyracoidea	family	1/1			
Hyracoidea	genus	1/3			
Hyracoidea	species	1/4			
Lagomorpha	family	2/2			
Lagomorpha	genus	5/12		-1.06	-0.95
Lagomorpha	species	12/86		-0.62	-1.88
Macroscelidea	family	1/1			
Macroscelidea	genus	4/4			
Macroscelidea	species	12/15		-1.3	-1.06
Microbiotheria	family	1/1			
Microbiotheria	genus	1/1			
Microbiotheria	species	1/1			
Monotremata	family	2/2			
Monotremata	genus	2/3		-0.72	-0.69
Monotremata	species	2/4		-0.97	-0.97
Notoryctemorphia	family	1/1			
Notoryctemorphia	genus	1/1			
Notoryctemorphia	species	0/2			
Paucituberculata	family	1/1			
Paucituberculata	genus	3/3			

Paucituberculata	species	5/5			
Peramelemorphia	family	2/2			
Peramelemorphia	genus	7/7			
Peramelemorphia	species	16/18		-0.13	0.97
Perissodactyla	family	3/3			
Perissodactyla	genus	6/6			
Perissodactyla	species	10/16		-0.07	-2.63
Pholidota	family	1/1			
Pholidota	genus	1/1			
Pholidota	species	4/8		1.18	0.94
Pilosa	family	4/5		1.87	2
Pilosa	genus	4/5		-0.96	0.36
Pilosa	**species**	**5/29**		**1.28**	**2.38***
Primates	family	15/15			
Primates	genus	48/68		-0.35	-1.33
Primates	species	64/351		-0.67	-1.27
Proboscidea	family	1/1			
Proboscidea	genus	2/2			
Proboscidea	species	2/3		-0.69	-0.69
Rodentia	family	18/32		0.66	-0.98
Rodentia	genus	82/450		-1.66	1.55
Rodentia	**species**	**90/2094**		**2.76***	**2.34***
Scandentia	family	2/2			
Scandentia	genus	2/5		-0.74	-0.74

Scandentia	species	3/20		-1.88	-0.84
Sirenia	family	2/2			
Sirenia	genus	2/2			
Sirenia	species	4/4			
Soricomorpha	family	3/4		-0.98	-0.99
Soricomorpha	**genus**	**19/43**		**7.11****	**2.59****
Soricomorpha	**species**	**21/392**		**10.65****	**3.56****
Tubulidentata	family	1/1			
Tubulidentata	genus	1/1			
Tubulidentata	species	1/1			
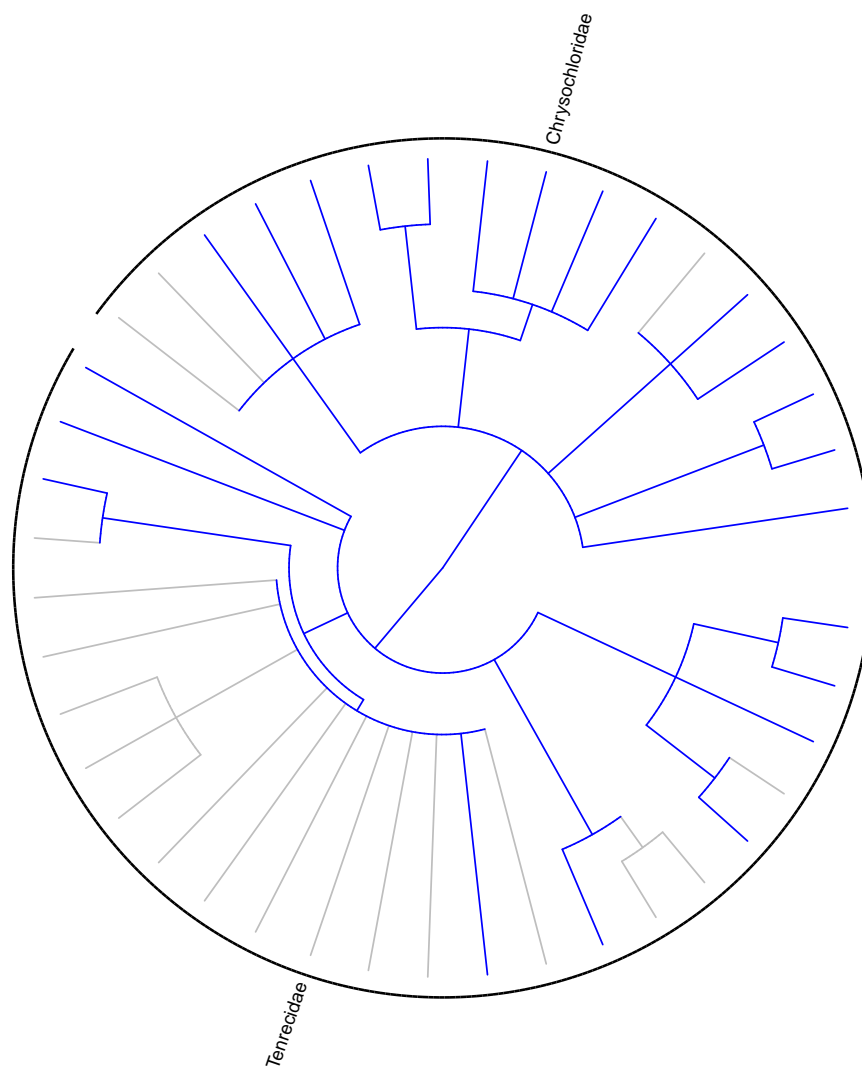
FIGURE A.1: Distribution of available morphological data across Afrosoricida. Edges are colored in grey when no morphological data is available or in blue when data is available.
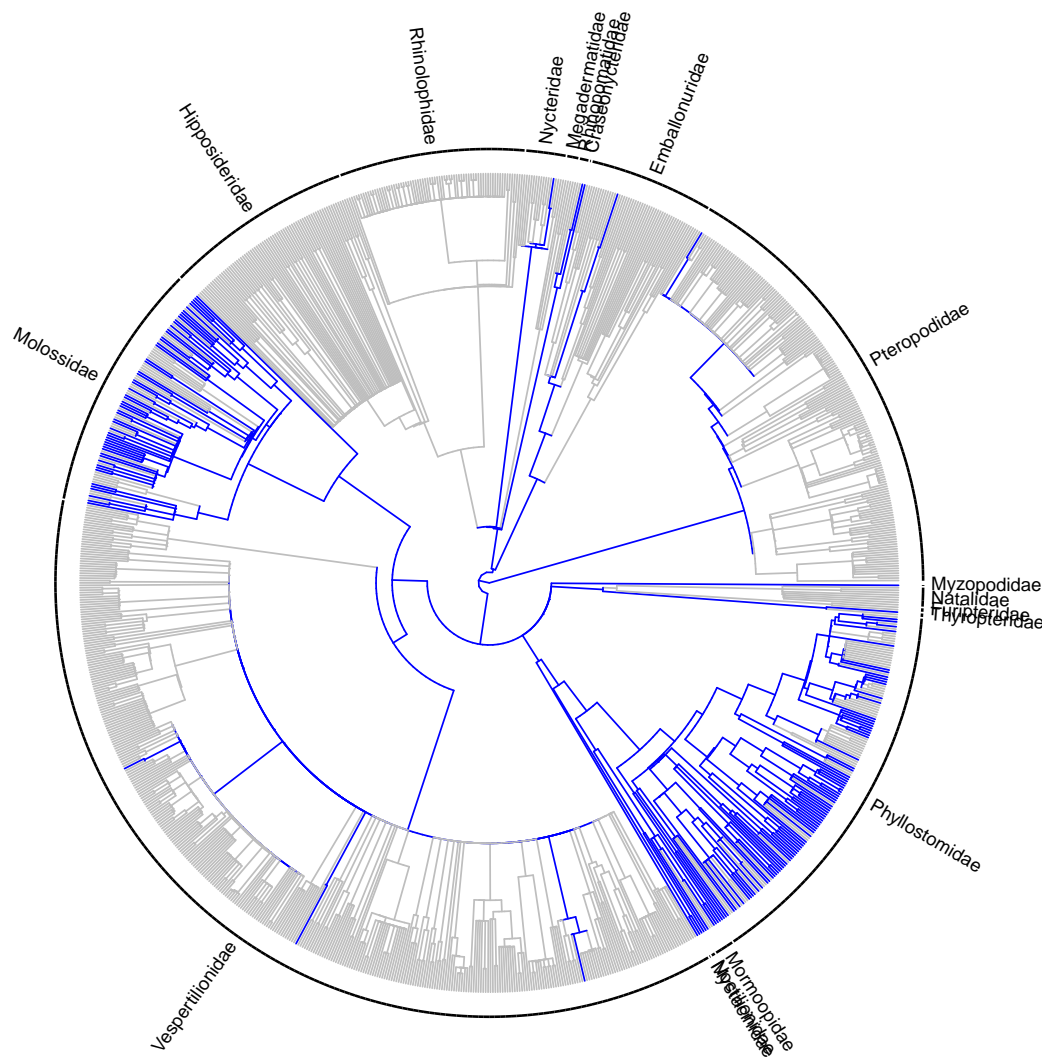
FIGURE A.2: Distribution of available morphological data across Chiroptera. Edges are colored in grey when no morphological data is available or in blue when data is available.
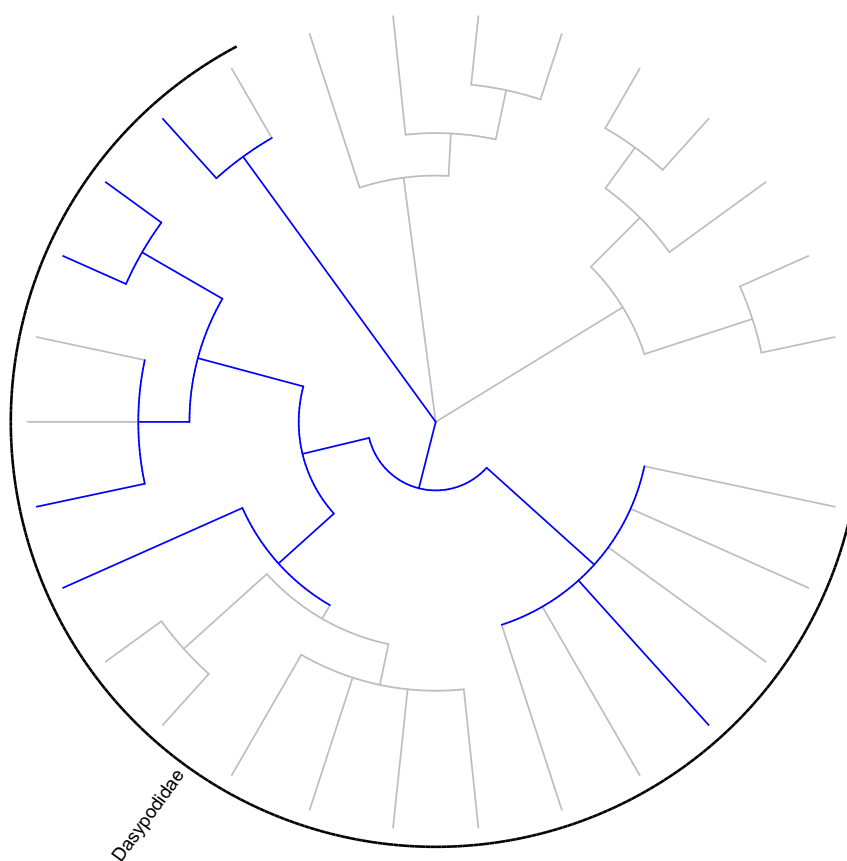
FIGURE A.3: Distribution of available morphological data across Cingulata. Edges are colored in grey when no morphological data is available or in blue when data is available.
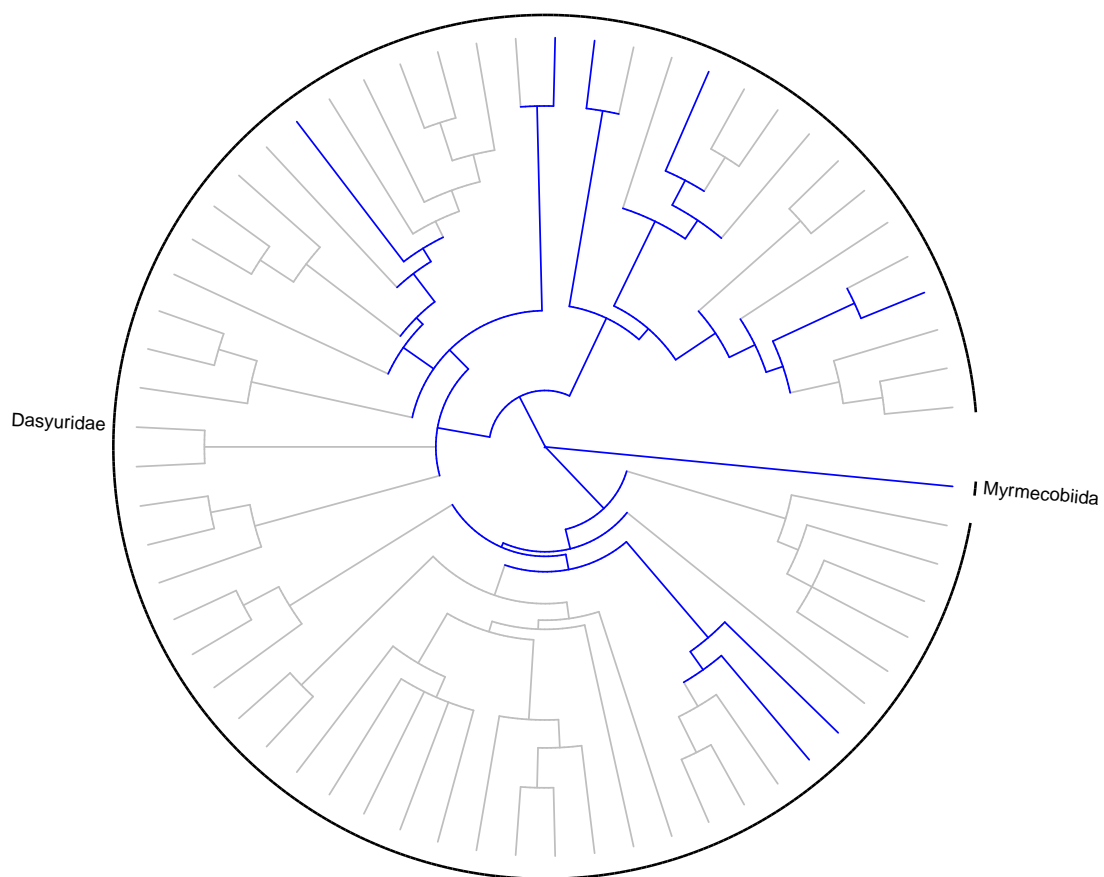
FIGURE A.4: Distribution of available morphological data across Dasyuromorphia. Edges are colored in grey when no morphological data is available or in blue when data is available.
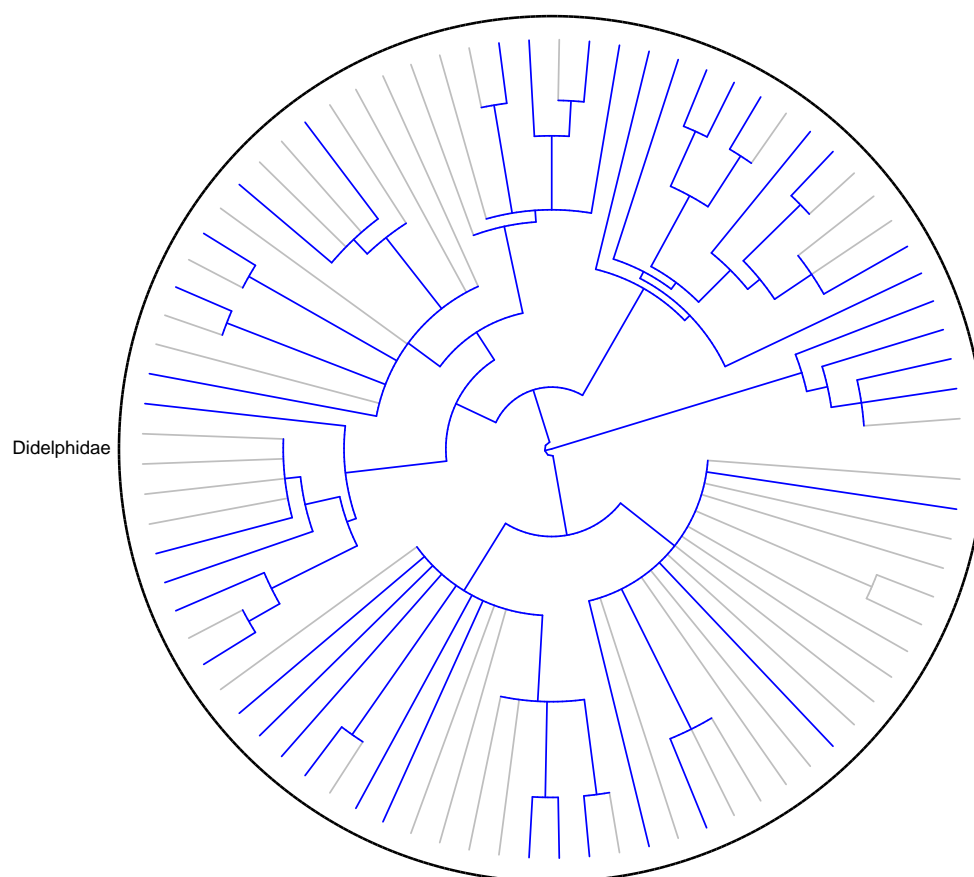
Didelphidae

FIGURE A.5: Distribution of available morphological data across Didelphimorphia. Edges are colored in grey when no morphological data is available or in blue when data is available.
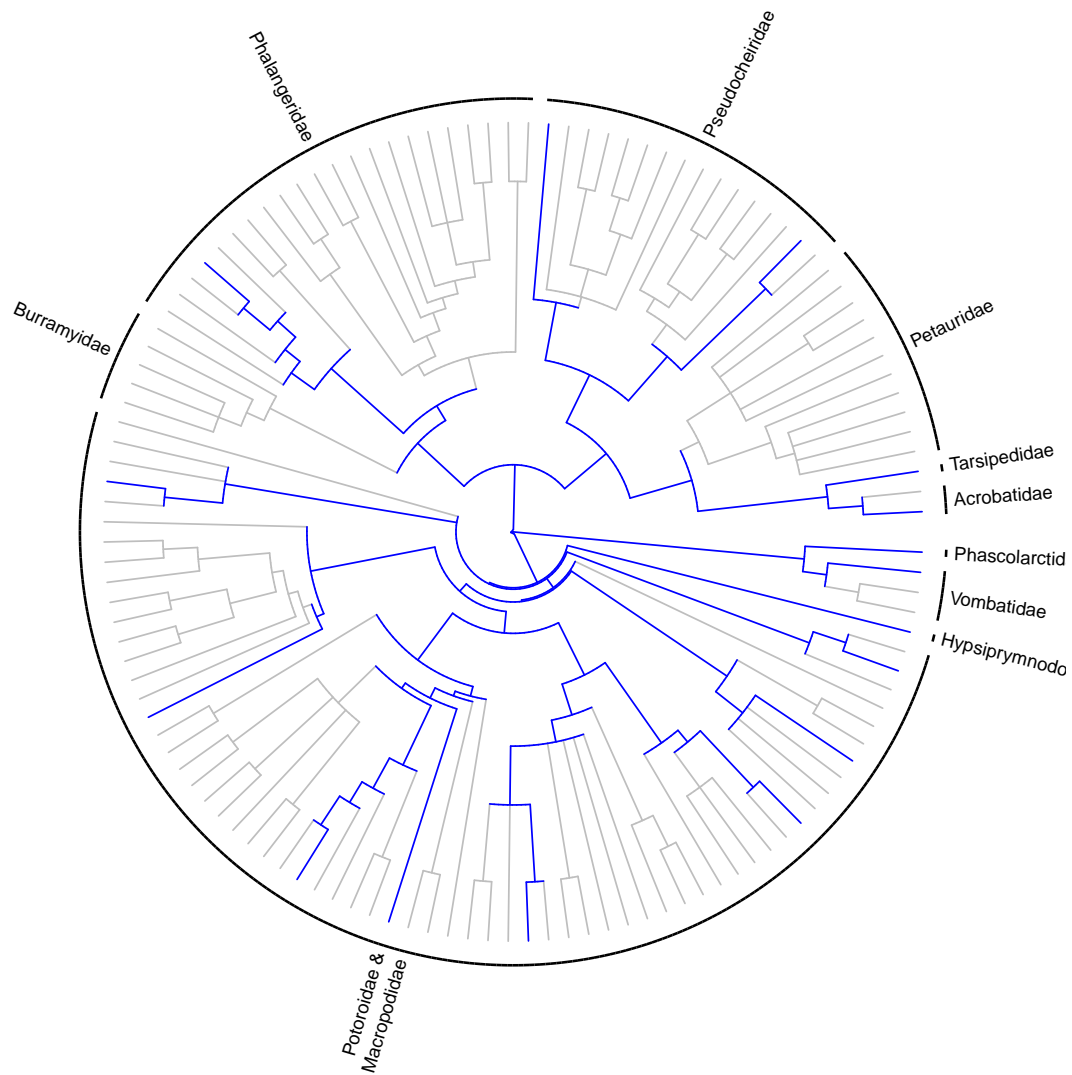
FIGURE A.6: Distribution of available morphological data across Diprotodontia. Edges are colored in grey when no morphological data is available or in blue when data is available.
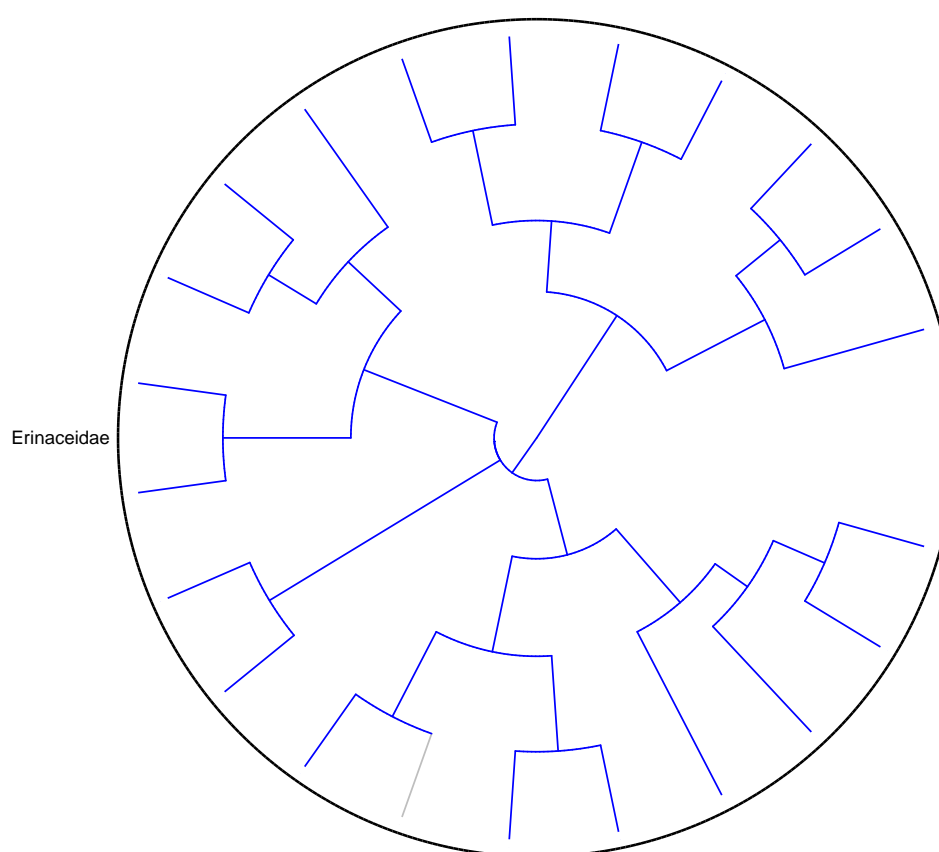
Erinaceidae

FIGURE A.7: Distribution of available morphological data across Erinaceomorpha. Edges are colored in grey when no morphological data is available or in blue when data is available.
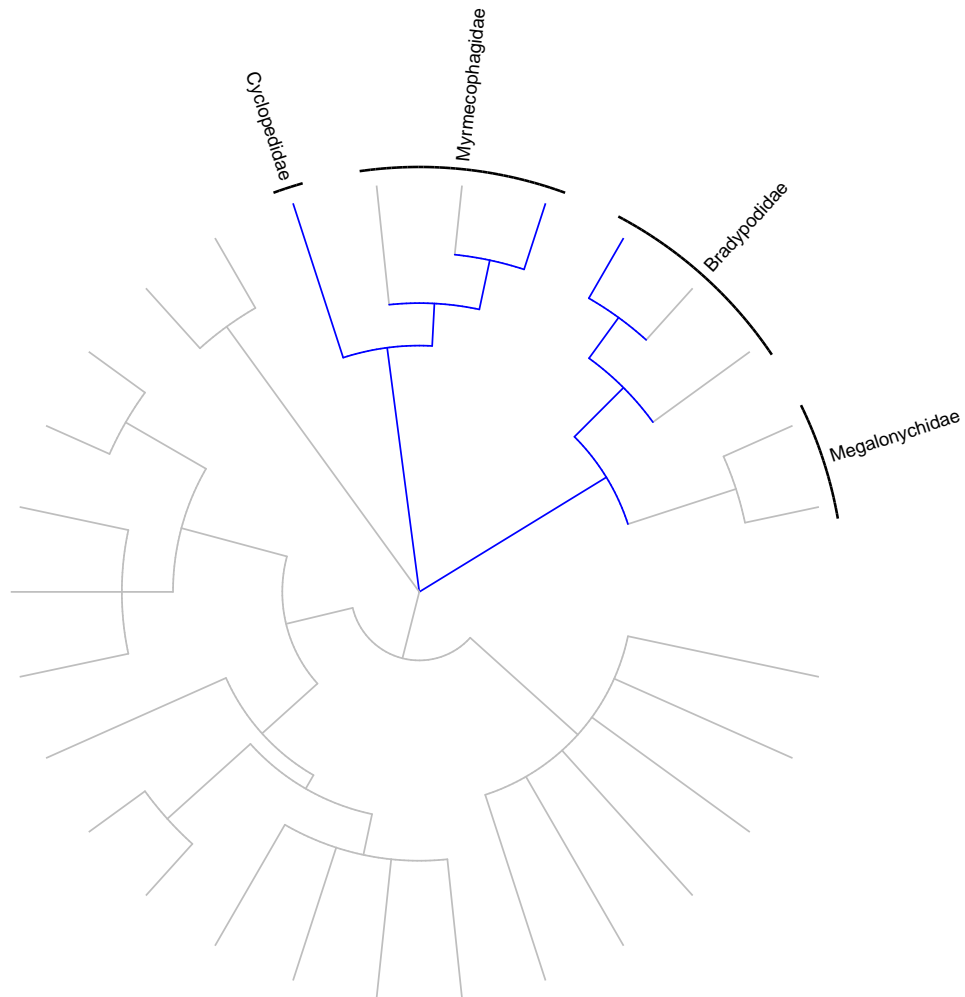
FIGURE A.8: Distribution of available morphological data across Pilosa. Edges are colored in grey when no morphological data is available or in blue when data is available.
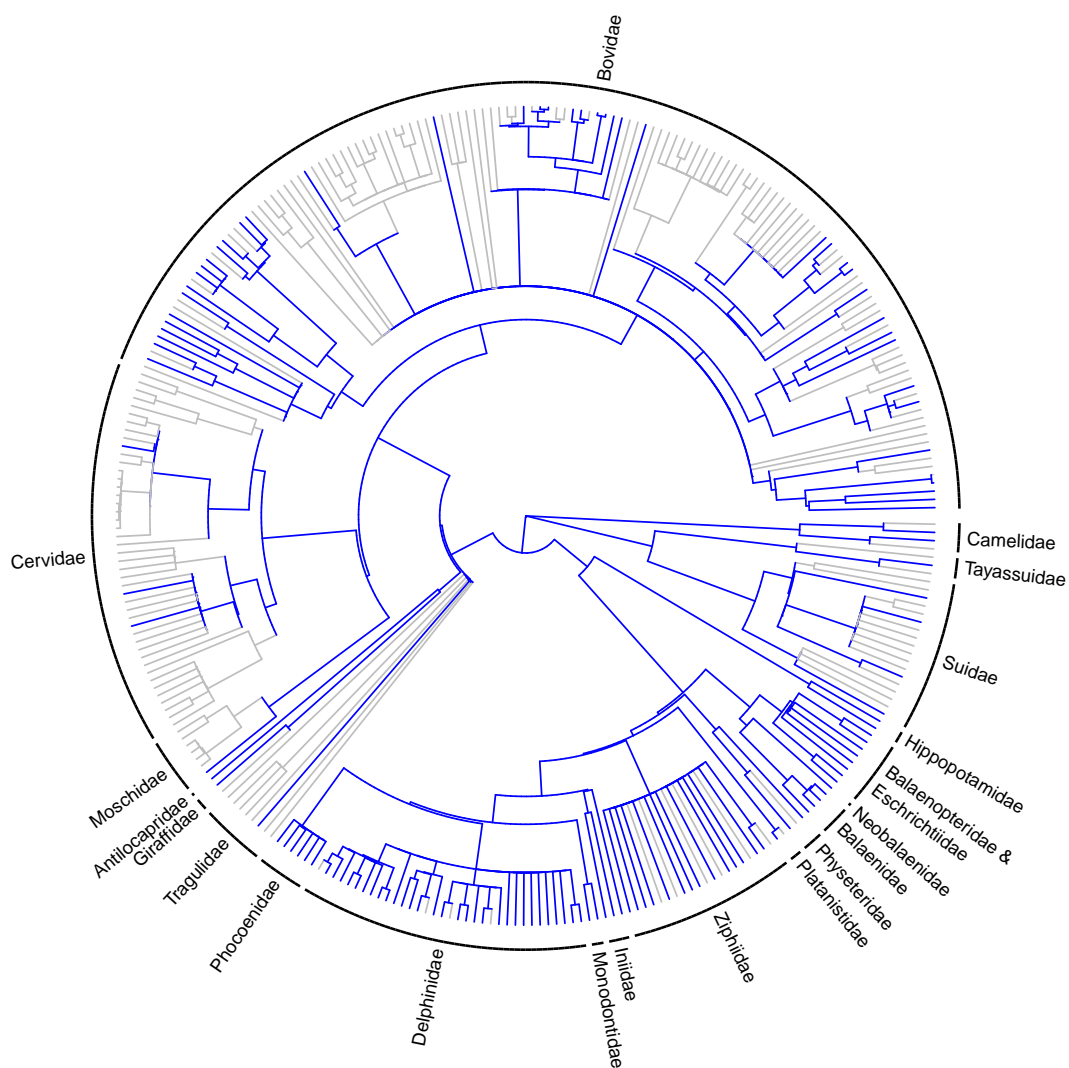
FIGURE A.9: Distribution of available morphological data across Cetartiodactyla. Edges are colored in grey when no morphological data is available or in blue when data is available.
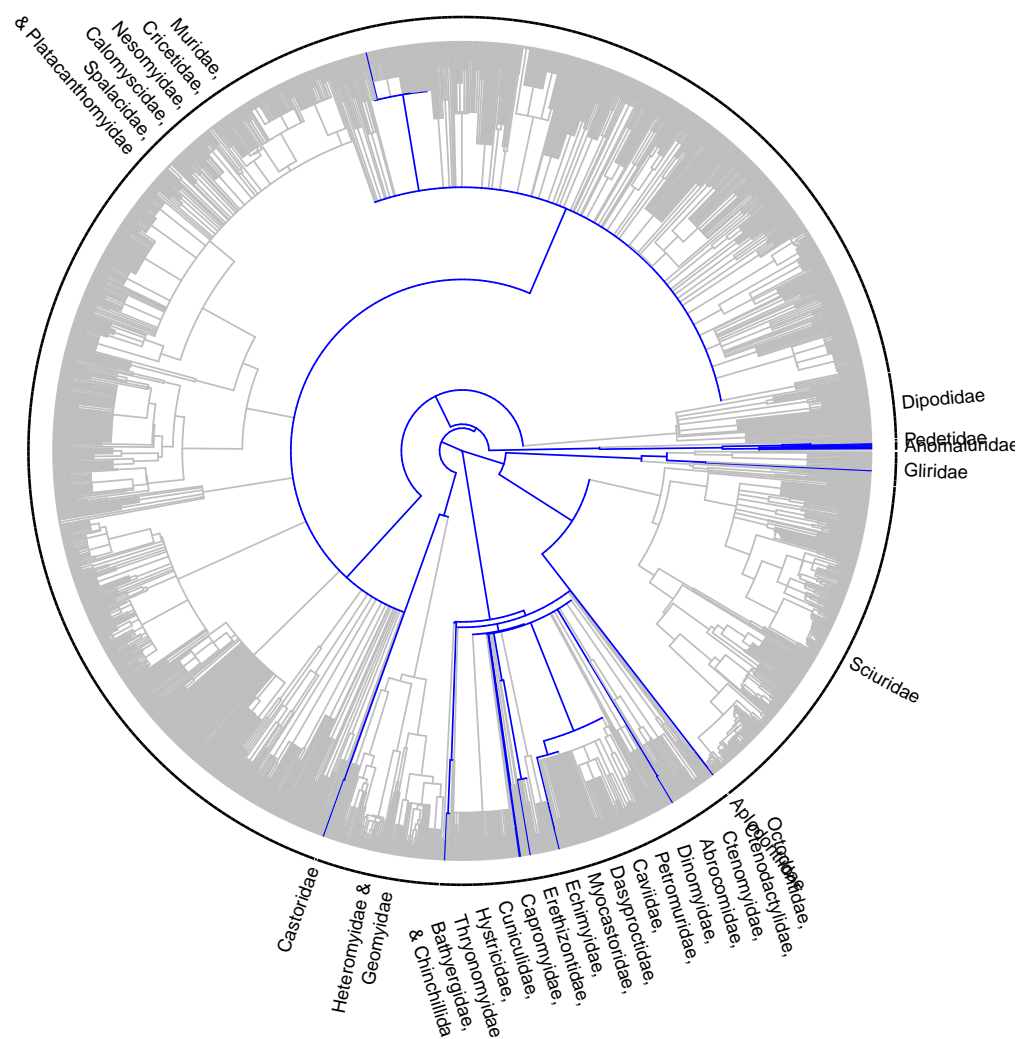
FIGURE A.10: Distribution of available morphological data across Rodentia. Edges are colored in grey when no morphological data is available or in blue when data is available.
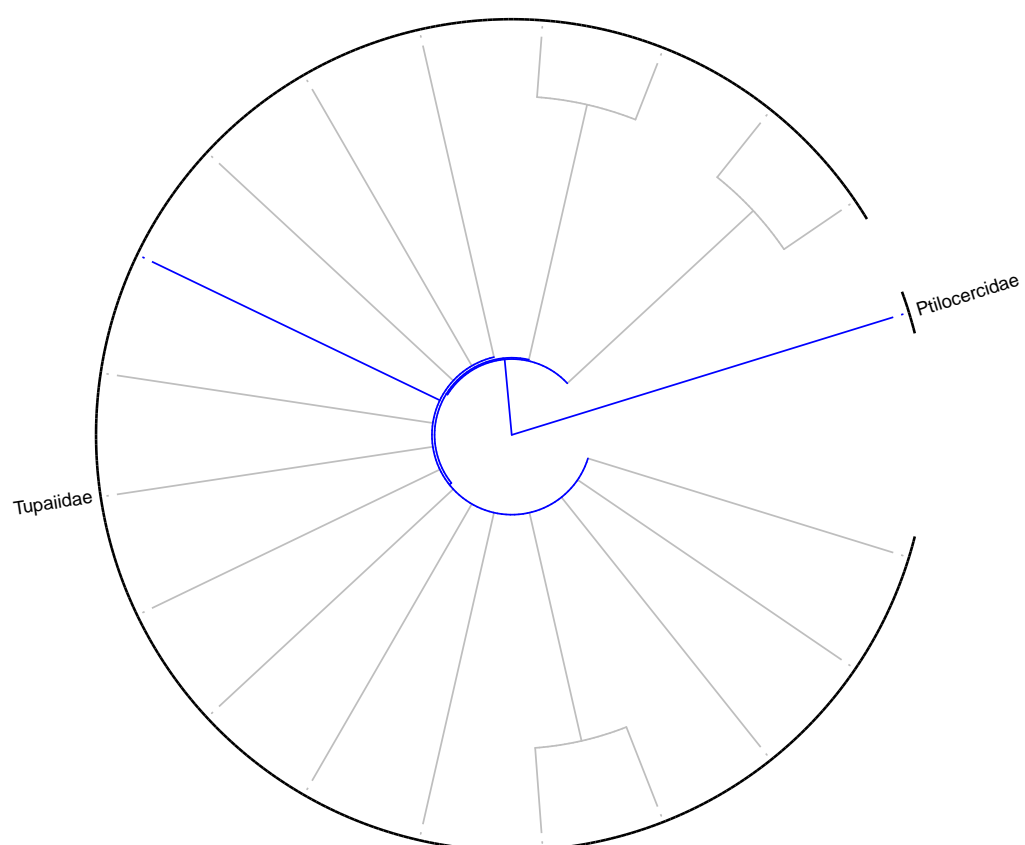
FIGURE A.11: Distribution of available morphological data across Scandentia. Edges are colored in grey when no morphological data is available or in blue when data is available.
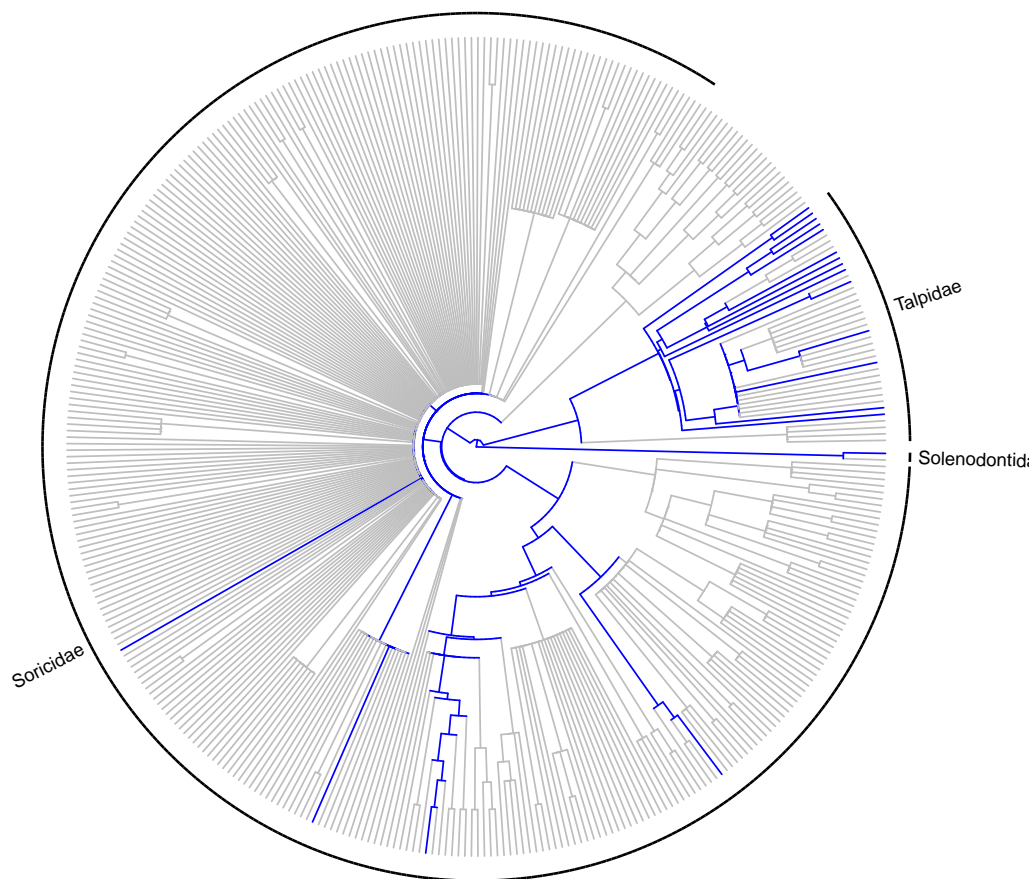
FIGURE A.12: Distribution of available morphological data across Soricomorpha. Edges are colored in grey when no morphological data is available or in blue when data is available.