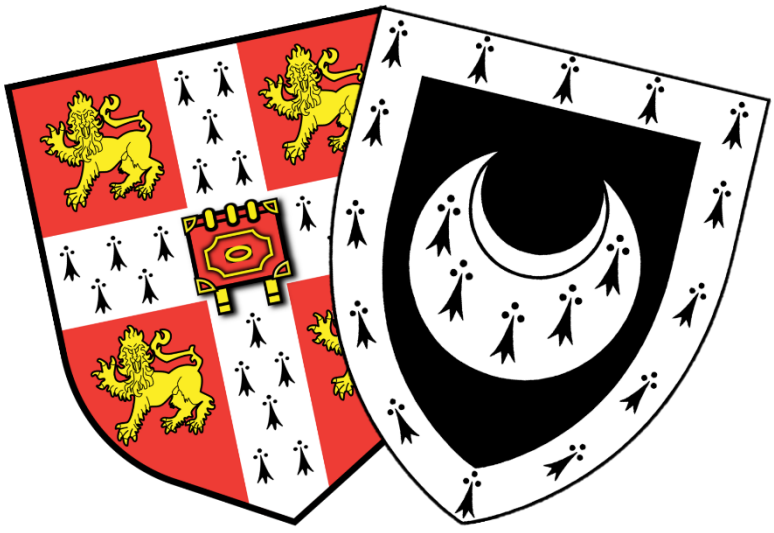
Assessing the information content of fossilizable data, with a focus on the Asian origin and diversification of Glires (Mammalia)

Aime Hall Rankin

Trinity Hall

December 2020



This thesis is submitted for the degree of Doctor of Philosophy at the University of Cambridge

For Tookay Rankin and Finny Rankin

Declaration

This thesis is the result of my own work and includes nothing which is the outcome of work done in collaboration except as declared in the preface and specified in the text.

It is not substantially the same as any work that has already been submitted before for any degree or other qualification except as declared in the preface and specified in the text.

It does not exceed the prescribed word limit for the School of Biology Degree Committee.

December 2020

Thesis summary

Assessing the information content of fossilizable data, with a focus on the Asian origin and diversification of Glires (Mammalia)

Aime Hall Rankin

Chapter One: introduction. My thesis explores how the information content of fossil Glires (rodents + rabbits) effects phylogenetic reconstruction and the implications for downstream analysis and interpretation. The first two chapters take a detailed look at the phylogenetic signal contained within fossil Glires, while the third and fourth chapters put interpretation of fossils into practice through analyses of phylogeny, biogeographic history, anatomy and geometric morphometrics.

Chapter Two: assessing the information content of the Glires fossil record. This chapter investigates the role of missing data and preservational biases in the ability of fossil Glires to accurately reconstruct phylogeny. In particular, I challenge the idea that fossils distort trees by occupying erroneously primitive positions. I use a method called 'artificial extinction', which degrades the information of a living species with known phylogenetic affinities to match the quality of data found within a real fossil. These 'artificial fossils' are then tested to see if they can recover the same phylogenetic positions in a tree as their living counterparts. My results indicate that these artificial fossil proxies of real fossil Glires can adequately reconstruct phylogeny. Tests on artificial fossils that are forced to sample characters evenly across morphological partitions suggest that the preservational biases within fossil Glires play a minor role in phylogenetic accuracy. Finally, I find no evidence that fossil Glires distort trees by means of a stemward slippage bias, further implying that fossil Glires are suitable for phylogenetic reconstruction.

Chapter Three: morphological data partitioning and phylogenetic signal in Glires. This smaller chapter carries on from Chapter Two's discussions on the importance of preservational biases within fossils. Some partitions are known to contain different phylogenetic signal to others, but little work has been done on whether a single partition contains differentiated signal across a tree. I investigate if morphological character partitions in Glires optimize as apomorphies in different parts of the tree. My results indicate that some morphological partitions resolve groups near the tips of a tree, whereas others are more likely to support groups near the root. I discuss the implication of this in phylogenetic analysis and the possibility of weighting schemes to take this information into account.

Chapter Four: origins, diversification and biogeography of Glires; the role of the Asian continent and uplift of the Tibetan Plateau. This chapter compares total-evidence approaches towards Glires phylogeny and how different topologies affect our interpretation of biogeographic history. Ultraconserved elements (UCEs) are known to provide unconventional hypotheses of basal rodent affinities and so I compare topologies made from fossil data and protein coding DNA to those that also include UCEs. I describe differences in the root of rodents, the positions of key fossils and the timing of rodent evolution. Glires are considered to be of Asian origin and so I also carry out biogeographic analyses on my two topologies. I find that standard total-evidence approaches fully support an Asian origin for the group, while UCEs provide a less resolved answer. I also compare some competing hypotheses on the origins of the three main rodent crown groups. Finally, I discuss how the timings and diversity data support a role of the Tibetan Plateau uplift in rodent diversification.

Chapter Five: anatomical sciuromorphy in 'protrogomorph' rodents. This chapter focuses on the interpretation of anatomical characters in fossil Glires and the implications for character evolution. I present a paper published from this thesis on anatomical sciuromorphy present within 'protrogomorph' rodents. I give a history of how the rodent master system has been interpreted from fossils over the years and describe the presence of derived musculature in a well-preserved Eocene rodent specimen. I describe the osteological features of this specimen, provide an identification and use geometric morphometrics to assess if its musculature is closer to modern day rodents or fossil taxa.

Chapter Six: general discussion and concluding remarks. I discuss the importance of my biogeographic, anatomical and phylogenetic results within the context of Glires evolution. I also recommend the inclusion of fossil data in total-evidence approaches where possible and give suggestions for future work.

Contributions

Chapter Two: I collected and analysed the data and Dr Rob Asher provided R scrips which I adapted to suit my dataset. For taxa not included in the Asher et al (2019) dataset, I selected the specimens, undertook microCT scans and coded the characters. I wrote the text and made the figures.

Chapter Three: I gathered and analysed the data. Dr Rob Asher provided R scrips which I adapted to suit my dataset. I made the figures and wrote the text.

Chapter Four: I gathered and analysed the data. I adapted a script provided by Bethany J. Allen (University of Leeds) to calculate squares estimates of diversity. I made the figures and wrote the text.

Chapter Five: This chapter is a reproduction of a paper published in the journal *Palaeontologia Electronica* (https://palaeo-electronica.org/content/2020/3043-protrogomorph-rodents). I have made some small adjustments to the paper so that the formatting is in keeping with the rest of the thesis. I wrote the paper before all co-authors, three reviewers and one editor commented on the work. Dr Bob Emry provided the specimen of *Ischyromys douglassi* and aided the interpretation of dental characters and identification. I analysed the data and made all but the following figures: Figure 24 (made by Dr Rob Asher), Figure 28 (photos taken by Dr Bob Emry, figure made by Dr Rob Asher), Figure 31 (made by Dr Rob Asher and myself), Figure 34 (made by Dr Rob Asher). Access to specimens and microCT data were made possible by several individuals: Lionel Hautier, Pierre-Henri Fabre and Monique Vianey Liaud (Montpellier), Brian Kraatz (Pomona), Jin Meng (New York), Ornella Bertrand and Mary Silcox (Toronto), Matthew Lowe, Natalie Jones and Ewan St. John Smith (Cambridge), Pip Brewer, Jerry Hooker, and Roberto Portela Miguez (London), and Jessica Nakano and Jennifer Strotman (Washington DC).

Acknowledgements

There are many to whom I wish to express my gratitude.

Firstly, I am immensely grateful to my supervisor Dr Rob Asher without whom this project would not have existed. His mentoring and support over the last four years have been invaluable. I also wish to thank my past supervisors Dr Anne Jungblut and Prof Rod Page for inspiring my love of evolutionary studies and encouraging me to pursue this PhD. I would also like thank my examiners, Dr Alex Liu and Prof Pierre-Henri Fabre for their insightful comments and suggestions.

All research requires the help of collaborators and colleagues, and my work has been no exception. I would like to extend thanks to all of the staff at the University Museum of Zoology for facilitating my access to specimens. I am also grateful to the curatorial teams at the AMNH, YPM, HCZ and ACZ who welcomed me into their institutions and gave me access to precious fossil material when I visited the USA. In particular, special thanks are owed to Haley Singleton (ACZ) who embarked on a road trip to Boston with me and the type specimen of *Franimys amherstensis*. There are also many individuals who shared specimens and microCT data with me for whom I would like to express my gratitude: Lionel Hautier, Pierre-Henri Fabre and Monique Vianey Liaud (Montpellier), Brian Kraatz (Pomona), Jin Meng (New York), Ornella Bertrand and Mary Silcox (Toronto) and Ewan St. John Smith (Cambridge), Pip Brewer, Jerry Hooker, and Roberto Portela Miguez (London), Mark Omura (Harvard) and Jessica Nakano and Jennifer Strotman (Washington DC).

I am equally indebted to those who passed on their skills to me. I am grateful to Phil Cox (York), Phil Morris (Florida) and Madeleine Geiger (Zurich) for introducing me to geometric morphometric methods. I must also thank Phil Cox for dissecting a capybara specimen with me and sharing his knowledge of rodent anatomy. I would also like to thank Joana Meier (Cambridge) for teaching me how to adapt my scripts for the computer cluster.

I would also like to thank the many specimens I have worked with. Their miraculous preservation and morphological curiosities have provided constant wonder and inspiration throughout the years.

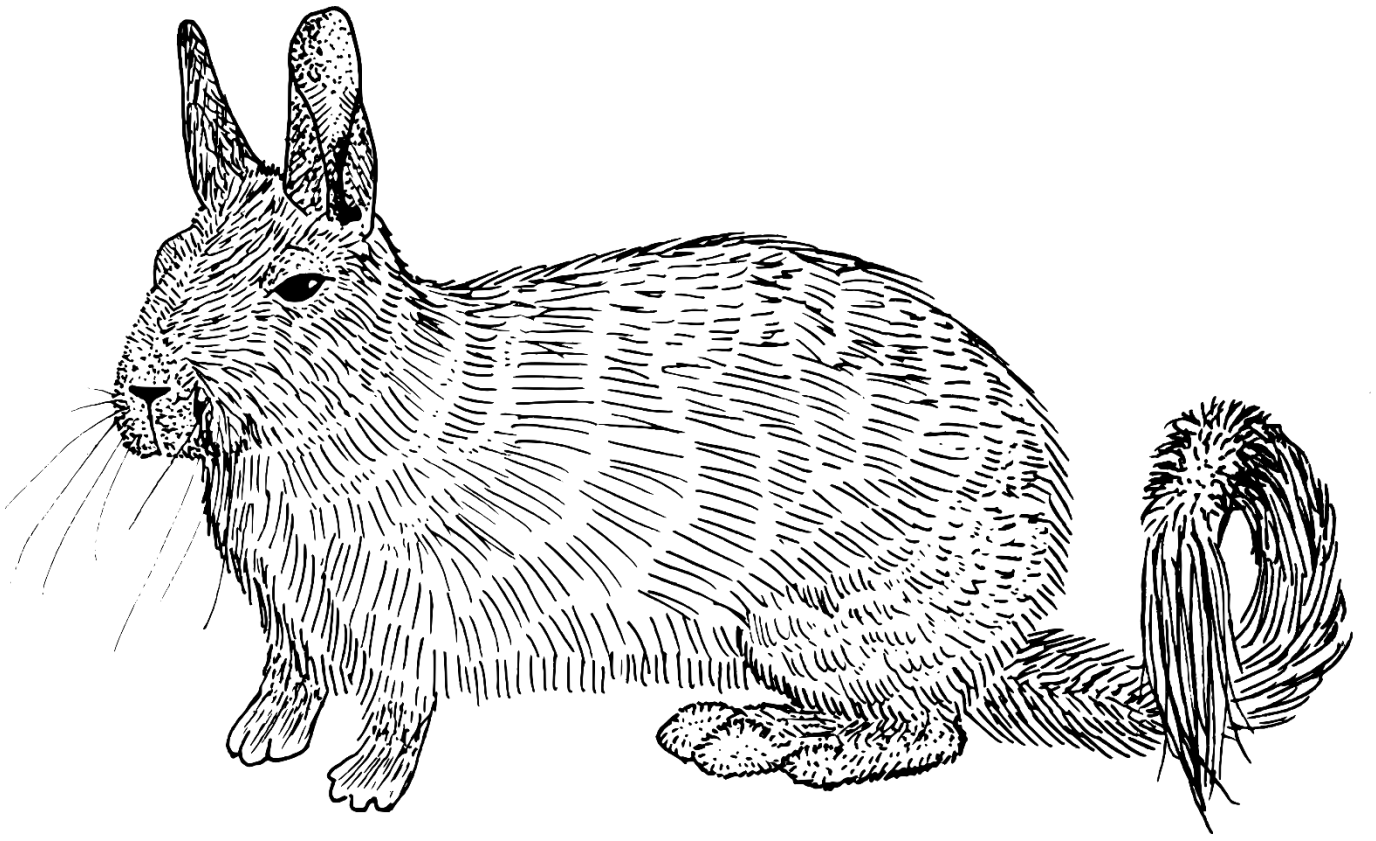
This research was generously funded by a Whitten Studentship from the Department of Zoology, for which I will be forever indebted. As well as providing the means for my study, I have gained many friends through the Whitten Studentship. Our cohort felt more like family and so I wish to thank them all for their support. In particular, I would like to show my appreciation for Claire Barnes, for her enduring encouragement, friendship, and for inspiring my interest in Asian zoology.

My time spent in the Austin Building at the Department of Zoology would not have been so enjoyable without the company of many fine mammologists. My thanks go to Rachel O'Meara, Madeleine Geiger, Jack McMinn, Calum McKay and Oscar Wilson for their wit, humour and support. I am also very lucky to have received friendship from members of the Trinity Hall MCR, and wish to express my gratitude to Trinity Hall and the University of Cambridge for providing a welcoming and social learning environment.

Finally, this thesis would not have been possible without the help, love and support from my partner Nicolas, my parents Caryl and Roger, my sisters and brother, and all of my family and friends back home - thank you one and all!

Contents

Chapter One: introduction



Lagidium viscacia

The fossil record and phylogenetic analysis

It is a widely acknowledged circumstance that the majority of species to have ever lived on Earth are now extinct. Our knowledge of these past life forms comes from the fossil record, which documents a wide variety of plants and animals over millions of years. By looking at this record of life, it is clear that the great wealth of species that we see around us today represents a drop in the ocean compared to the diversity of taxa that once lived (Simpson, 1952). In order to understand the evolutionary history of life on Earth, it is often necessary to refer back to these extinct forms. Fossils are usually the only physical evidence we have of extinct species and of macroevolutionary processes, which makes them an invaluable source of data.

Including fossils in phylogenetic analyses can have many benefits. The foremost of course, is an increased density of taxon sampling, which could not be achieved by molecular data alone. In addition, the characters states provided by fossil data can alter our hypotheses of character polarity and homology (e.g., Smith, 1998; Asher et al., 2008; Edgecombe, 2010). Fossils can also be used to infer minimum dates, provide calibration for molecular clocks (dos Reis et al., 2016), estimate divergence times, break up long branches (Giles et al., 2017) and infer ancestral character states and sequences of character evolution (Puttick, 2016). Hypotheses of biogeography are also enhanced by fossil data (Marivaux and Boivin, 2019). There is no doubt that fossils have the potential to provide significant contributions to phylogenetic analyses. However, much of the results are contingent on where fossils are placed on the tree and how their anatomy is interpreted. Given the minimum age of a fossil, where it is positioned in relation to other taxa on a tree could adjust the age of a clade by millions of years (Doyle and Donoghue, 1993; Magallón and Sanderson, 2001; Magallón, 2004). This in turn could alter our understanding of where and under what environmental pressures a clade diversified. Given a fossil's set of morphological character states, where it is placed on a tree could also lead to different theories of how anatomical adaptations evolved. The accurate placement of fossil taxa is therefore an important consideration of any phylogenetic study which uses them.

Phylogenetic accuracy and fossil placement are often discussed under the context of missing data. The process of fossilization results in an inevitable degradation of material and so fossils represent incomplete data. Soft tissues are the first body parts to decompose after death and DNA is surprisingly short lived: the oldest translated amino acid sequences are only 1.77 Ma old (Cappellini et al., 2019). Soft tissues can of course preserve under exceptional conditions, such as in Lagerstätten, but this is not a common occurrence and most fossils comprise of osteological data only (Allison, 1988). The fossil record is therefore subject to preservational biases towards hard tissue. Nevertheless, fossils remain our best evidence of extinct creatures.

For most of the 19th and 20th centuries, fossil material was thought to be essential for understanding the relationships between living taxa (Smith, 1998). However, this viewpoint started to change in the late 1970s with the onset of cladistic analysis. Researchers began to debate about the limitations of fossils, given their missing data, and how best to incorporate fossils into analyses of phylogeny. The harshest criticisms of fossil data came from palaeontologists themselves, such as Collin Patterson. Patterson (1981) argued that fossils do not contain enough information to accurately place them on the tree of life. Further studies claimed that the level of missing data in fossils rendered them either uninformative for phylogenetic analysis or at worst, misleading (Hennig et al., 1981; Ax, 1987). Both Hennig et al. (1981) and Ax (1987) advocated that fossils, if included at all, should only be allocated to stem groups after a tree of living taxa is made. However, not all researchers at the time felt this way. Gauthier et al. (1988) performed experiments which showed that fossils were capable of influencing the relationships between living taxa, and so they contended that fossils did indeed contain useful phylogenetic information. Researchers such as Donoghue et al. (1989) encouraged the incorporation of fossil data wherever possible. Yet the debate of whether or not to include fossil data has not gone away. Recent advances in molecular data analysis have made molecular based phylogenies the standard in inferring relationships, with some authors reasoning that morphological data have been superseded (Scotland et al., 2003) or that they should not be used to infer phylogeny altogether (Halanych, 2016).

In light of continuing discussion on the efficacy of fossil data, great effort has been made to find empirical ways of testing the 'reliability' of fossil data. Early simulation studies by John Wiens for example (Weins, 1998; Wiens, 2003a,b; Wiens, 2005), tested the effects of increasing missing data against the addition of more morphological characters in phylogenetic analysis. Wiens found that in general, large amounts of missing data have little effect as long as character sampling is sufficiently thorough. As well as simulated data, research has also been carried out on real morphological characters. Taphonomic studies of living animals have been used to understand the nature of decay and to assess the phylogenetic signal of characters that are most likely to preserve. For example, Sansom et al. (2010) examined the preservational bias towards soft tissues in *Amphioxus* and ammocoetes. They found that the morphological characters that were most likely to preserve were also the plesiomorphic ones. This, they hypothesised, would lead to fossils occupying erroneously primitive positions in phylogeny.

Another approach to investigating the phylogenetic information content of fossils is to artificially degrade the characters of living species and see how this would affect congruence to a known topology. In 2013, Sansom and Wills compared the deletion of soft tissue characters in a taxon to the deletion of an identical number of characters spread randomly across hard and soft tissue data

partitions. They found that trees where the soft tissue data had been deleted were less congruent to the optimal tree than when they had deleted characters at random. Furthermore, they found that when soft tissues were deleted, a taxon was more likely to experience stemward slippage rather than crownward slippage. It should be noted however, that some of their optimal trees were based on small molecular datasets that no longer reflect current ideas of animal relationships. Another study, conducted by Springer et al. (2007), degraded the information of mammal species. Springer et al. replaced whole orders of mammals in their matrix with artificially degraded taxa and then tested congruence with their optimal tree. Under a third of their artificially fossilized mammal orders were able to return the same phylogenetic positions as in the optimal tree.

Finally, there is a method which has been developed that is based on the missing data content of real fossils. In artificial extinction, the information provided by a living species is reduced to only the characters that are coded in any given fossil specimen. For example, a character coded as '?' in a fossil would then be coded as '?' in the living species. The 'artificial fossil' then replaces its living species counterpart in a phylogenetic analysis. If the artificial fossil recovers the same phylogenetic relationship as we know the living species to do, then we can assume that the information content of the fossil template is sufficient enough to accurately place the real fossil on the tree. This method was first attempted by Asher and Hofreiter (2006) who used the fossil *Parageogale* as a template to make artificial fossils of living tenrecids. Their topologies that contained an artificial fossil were not significantly incongruent with their well-corroborated tree of tenrecids, suggesting that *Parageogale* is capable of accurately reconstructing phylogeny. A further study by Pattinson et al. (2015) performed artificial extinction experiments on 85 fossil primates. Their results suggested, at least for primates, that fossils are capable of accurately reconstructing known phylogeny.

Another element of missing data in fossils that has been investigated is the bias towards preservation of denser osteological material (e.g. teeth) over softer ones. It has been shown that different parts of the body contain different phylogenetic signal (Pattinson et al., 2015; Mounce et al., 2016; Sansom et al., 2017). The morphological characters of an organism can be split into different parts called 'morphological partitions' and often these partitions are based on preconceived notions of function (e.g. crania and postcrania). Fossils that sample across these partitions are better equipped to accurately construct phylogeny (Pattinson et al., 2015; Mounce et al., 2016). So as well as the volume of missing data, the representation of different morphological partitions can have a large effect on the performance of fossils in phylogenetic analysis.

Glires: the most successful mammals on Earth

Glires (Rodentia + Lagomorpha) act as a good study group for artificial extinction and other experiments due to decent knowledge of their phylogenetic relationships, their relatively complete fossil record and large sample size. Glires comfortably hold the title of the world's most successful mammal group. They fulfil this role through a number of different criteria. They are the most speciose group of mammals, with diversity estimated to range from 2,277 species to 2,600 species (Wilson and Reeder, 2005; Fabre et al., 2015; Wilson et al., 2016; Elia et al., 2019). The majority of this diversity is accounted for by Rodentia, the larger of the two groups, comprising over 95% of glires species. As well as accounting for 30–40% of all living mammal diversity, Glires have a global distribution. Glires are native to every continent except Antarctica, although rodents have made it as far as South Georgia with our unintended help (Martin and Richardson, 2017). As well as being found in most places, Glires practice a number of different lifestyles, from scansorial to arboreal and semiaquatic to fossorial. They have evolved several different locomotion styles, which allow them to move around their environment, including gliding and ricocheting. Glires are also represented by a wide variety of diets including herbivory, granivory, omnivory, carnivory, fruigivory and vermivory (a specialised diet of worms). Finally, extant Glires range in size from as little as ca. 3g to over 60kg, with the largest ever rodent, the extinct *Josephoartigasia monei*, weighing in around 1000kg (Millien, 2008; Rinderknecht and Blanco 2008).

Specialised masticatory systems and systematics

One thing all Glires have in common is a highly specialised dentition for gnawing food. Rodents have an upper and a lower pair of ever-growing incisors, while lagomorphs have two pairs. The incisors are followed by a large diastema, molars and premolars, which are variably reduced in number. Continuously growing incisors are not unique to Glires as diprotodonty exists in the aye-aye, some hyrax species, the common wombat, the extinct multituberculate *Ptilodus* (Krause, 1982) and the extinct marsupial *Diprotodon* (Owen, 1870). However, the highly specialised masticatory muscles of Glires are more unusual and are another key component to their success. Four arrangements of the masseter muscle are observed in Glires: protrogomorphy, sciuromorphy, hystricomorphy and myomorphy. The masseter muscle in rodents is divided into several slips, which includes (in order of depth) the superficial masseter, deep masseter and zygomaticomandibularis. Where the different parts of the masseter attach to the cranium varies across rodents. In protrogomorphy, the masseter muscle's attachment is limited to the zygoma, as it is in many other mammal species (Figure 1). This anatomical condition is seen in Lagomorpha, the mountain beaver (*Aplodontia rufa*), some bathyergid moles (Cox and Faulkes, 2014), but most predominantly in groups of fossil Glires. In sciuromorphy, the

deep masseter extends forwards off the zygomatic arch and attaches onto the rostrum (Figure 1). This is coupled with a widening of the anterior root of the zygoma to form a structure known as the 'zygomatic plate'. Sciuromorphy is found in squirrels, beavers and Geomyoidea. In hystricomorphy, the deep masseter is restricted to the zygoma, but the zygomaticomandibularis extends forwards and invades an enlarged infraorbital foramen before attaching to the rostrum (Figure 1). This arrangement is found in anomalurids, dipodids, caviids, porcupines and some dormice. The last condition, myomorphy, combines the deep masseter extension and zygomatic plate of sciuromorphy and the zygomaticomandibularis extension and infraorbital invasion of hystricomorphy (Figure 1). This arrangement appears in murids and some dormice.

It is now generally accepted that Lagomorpha and Rodentia are sister groups, with the latter being split into three main clades: Sciuromorpha, Ctenohystrica and Myomorpha (Figure 1; Fabre et al., 2012; Asher et al., 2019). However, ideas of rodent taxonomy in the past were heavily influenced by the complicated system of masticatory muscles, and it took a while to arrive at the modern classification we recognise today. In 1855, Brandt divided the rodents into three groups based on the different conditions of the masseter muscle: Sciuromorpha, Hystricomorpha and Myomorpha. A few years later, Zittel (1893) formally defined 'Protrogomorpha', a group of rodents based on anatomical protrogomorphy. This group contained all of the fossil rodents as well as *Aplodontia*. Molecular based studies now show that these groupings based on anatomical definitions are unnatural and that the arrangements of the masseter muscle have evolved independently in several rodent groups (Hartenberger 1985; Fabre et al., 2012; Asher et al., 2019; Swanson et al., 2019). For example, the modern grouping of Sciuromorpha has representatives of all four conditions of the masseter muscle.

Later on in 1899, Tullberg conceived a new way to group rodents based on the lower jaw. Rodent lower jaws can be split into two categories (Figure 1). The angular process can either be in line with the plane of the incisor alveolus (defining Tulberg's group Sciurognathi) or at an angle (as in Tullberg's Hystricognathi). Tullberg's Hystricognathi is still recognised as a monophyletic group today within Ctenohystrica, but rodents with sciurognathous jaws comprise the rest of Rodentia (Figure 1). Another system to abandon the idea of masseteric classification was developed by Landry (1999). In Landry's new scheme, rodents were split according to the position of the nasolacrimal canal in relation to the incisor alveolus. His Entodacrya group possessed a canal running internally all the way down to the posterior end of the incisor alveolus, which allied the ctenodactylids and Hystricognathi. This group is still recognised as monophyletic today, but under the more frequently used name 'Ctenohystrica'

(Huchon, 2000). His Sciurognathi group, where the canal is partially exposed, is no longer recognised as a natural grouping.

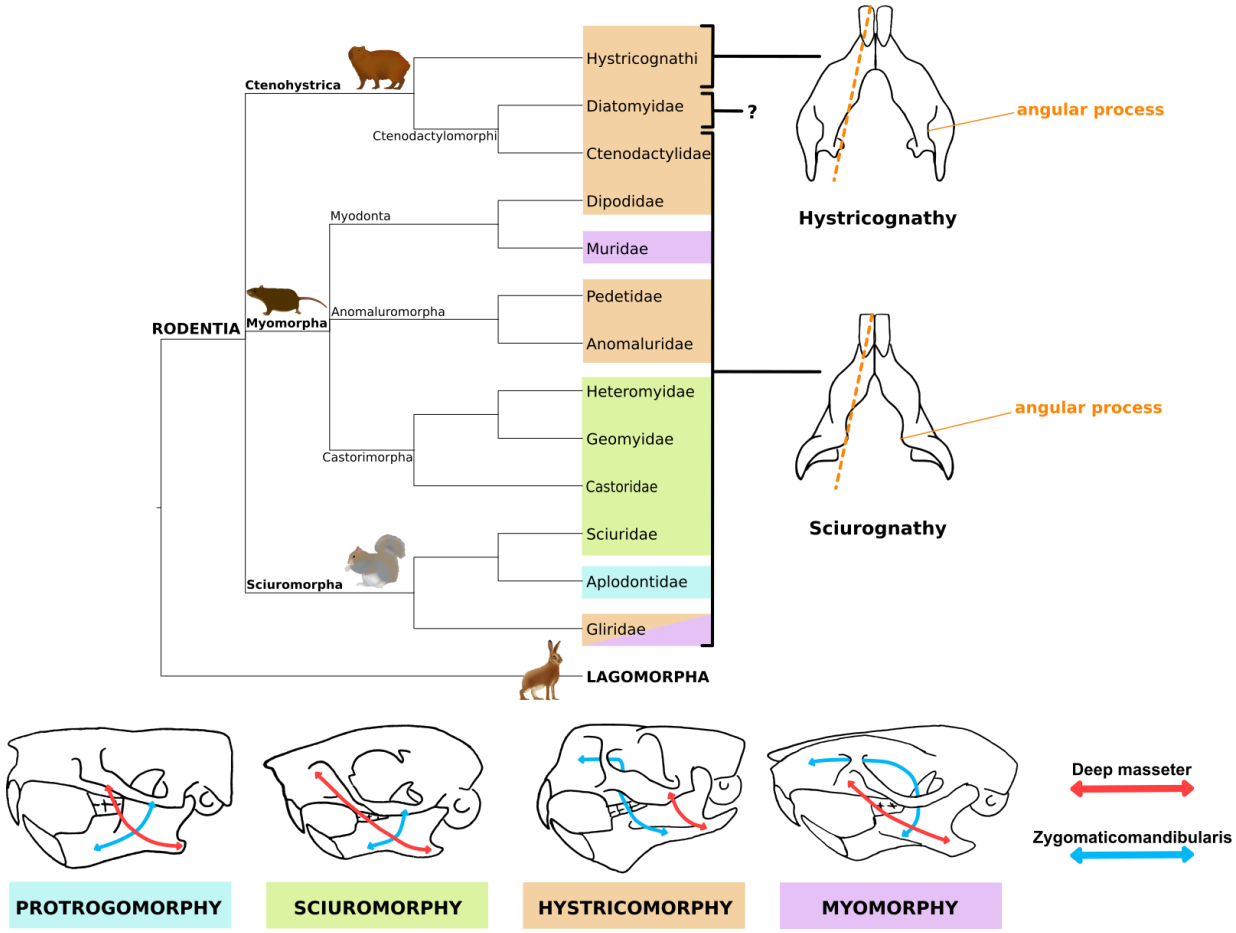


Figure 1: Rodent masseter muscle arrangements and angular process positions mapped onto a simplified topology of Glires. The uncertain angular jaw position of Diatomyidae reflects the intermediary condition in *Laonastes* described by Hautier et al. (2011). The tree is based on Fabre et al. (2012), Asher et al. (2019) and Elia et al. (2019). Figure layout is based on Hautier et al. (2015).

While the masseteric system is no longer used for classification purposes, the concept of a 'protrogomorph' group of animals has persisted, with the term 'protrogomorph' sometimes referring to members of a clade (Catzeflis et. al., 1992; Qiu and Li, 2005; Li et al., 2015), to members of a grade (Wood, 1965; Dawson, 1977; Vianey-Liaud; Korth, 1994), or describing the anatomical condition (Heaton 1966) or sometimes a grade and an anatomical condition interchangeably (Wood, 1980). The majority of the earliest fossil rodents have an anatomically protrogomorphous skull (Wood, 1962) and ancestral state reconstructions performed on recent topologies suggest that the ancestors of Sciuromorpha, Myomorpha and Rodentia itself were anatomically protrogomorphous (Swanson et al., 2019). However, the derived anatomical condition of sciuromorphy may be much older than first suspected. The fossil group 'Ischyromyidae' is often regarded as an exclusively protrogomorphous group and comprises of a paraphyletic assembly of several North American rodents from the

Paleocene and Eocene: *Franimys*, *Ischyromys* and *Paramys* to name a few (Korth, 1994). In 1976, the eminent anatomist A.E. Wood tried to convince the scientific community that some specimens of *Ischyromys* (which he referred to as *Titanotheriomys*) were anatomically sciuromorphous. This was met with criticism, most notably from his former student C.C. Black (1968), but while their debate was never settled, more recent studies still refer to this group of rodents as anatomically protrogomorphous (Korth, 1994). However, such discussions are worth revisiting in order to fully understand the evolution of the masseter system in rodents. For example, the application of geometric morphometric techniques to fossil data and new evidence of sciuromorphy in *Ischyromys* is discussed by Rankin et al. (2020). The results of this paper, which are presented in Chapter Five of this thesis, highlight the mosaic nature of masseter evolution and indicate that not all early-rodent clades were anatomically protrogomorphous.

Origins of Glires in Asia

Like in many other mammal groups, the timing and setting of the origin of Glires is of great interest and has led to several discussions on the matter. Glires has good molecular sampling and a fairly well sampled fossil record, but despite uncovering much of their evolutionary history throughout the years, there are still parts of their past which lack resolution.

Currently, molecular based estimates suggest that Glires appeared sometime between the Late Cretaceous (Hallstrom and Janke, 2010; dos Reis et al., 2012; Tarver et al., 2016) and the early Cenozoic (Douzery et al., 2003; Kitazoe et al., 2007; Wu et al., 2012). The oldest recognisable fossil Glires date to the Early Paleocene, just proceeding the K-Pg extinction event. These ancient fossils are so far only known from Asia and are allocated to one of two groups: eurymylids and mimotonids (Li, 1977; Li et al., 2016). Sometimes named collectively as ‘Mixodontia’, these animals are thought to possess the ancestral characters of Glires. Mimotonids have two pairs of upper and lower ever-growing incisors, the defining feature of grade ‘Duplicidentata’, and so are considered to be more closely related to Lagomorpha than Rodentia (Sych, 1971). Eurymylids only have one pair of upper and lower ever-growing incisors, characters of the grade ‘Simplicidentata’, and so are thought to be more closely related to Rodentia. Mimotonids can be referred to as ‘non-lagomorph duplicidentates’, and eurymylids as ‘non-rodent simplicidentates’ (Dashzeveg and Russell, 1988). In reality, the division between eurymylids and mimotonids is oversimplified and they do not constitute a monophyletic group (Meng and Wyss, 2001; Meng et al., 2003; Asher et al., 2005).

While an Asian origin of Glires seems very likely given the distribution of eurymylids and mimotonids, the origins of the three rodent crown groups (Sciuromorpha, Myomorpha and Ctenohystrica) is less certain. There are plenty of basal fossil Glires in North America, such as the rodentiaform family Algomyidae and the slightly younger 'Ischyromyidae', which have been used to suggest possible North American origins for Rodentia (Sloan, 1969; Wood, 1977). For example, the rodentiaform *Alagomys* is found on both sides of the Pacific, but current age correlates indicate that North American specimens from the Late Paleocene (Late Clarkforkian NALMA) predate their Asian neighbours from the Early Eocene (Bumbanian ALMA) by millions of years (Dashzeveg, 1990; Tong and Dawson, 1995; Bowen et al., 2012). Ischyromyidae are exclusively North American and latest Paleocene members such as *Acritoparamys* and *Paramys* were contemporaneous with the later Asian rodentiaformes and eurymylids (Korth, 1994; Figure 2). The contrast between the distribution of ancient ischyromyid and alagomyid fossils in North America, with their presumed ancestors in Asia, has been termed the 'Paleocene Paradox' (Dawson, 2015; Fostowicz-Frelik, 2020).

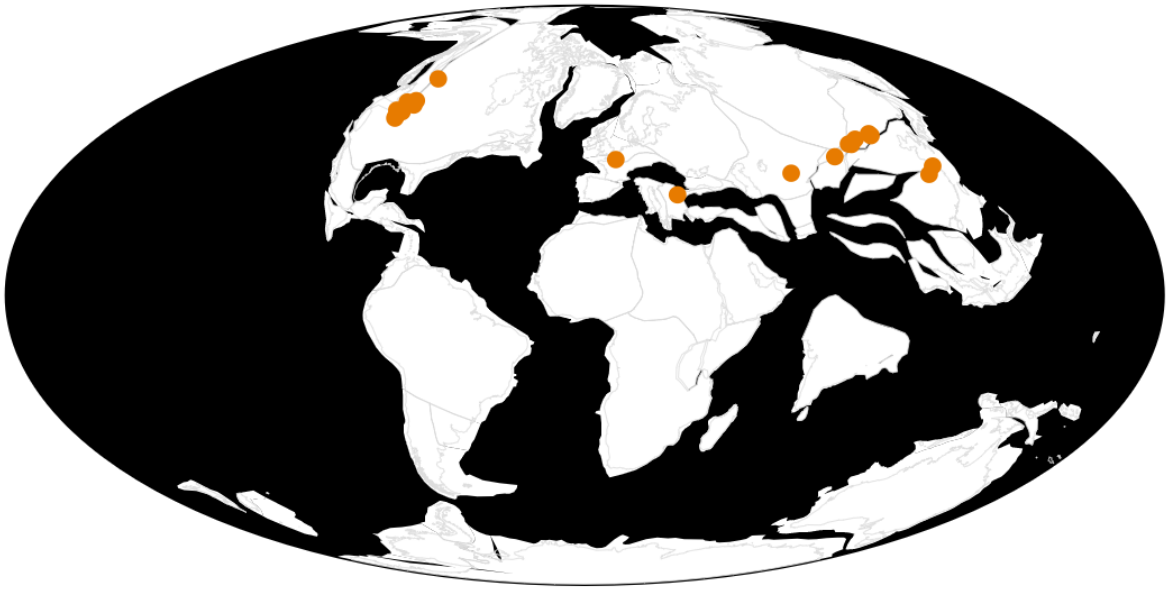


Figure 2: Distribution of fossil Glires during the Paleocene. Orange circles represent occurrence records. Data and map sourced from PBDB (www.paleobiodb.org).

The phylogenetic affinities of 'ischyromyid' rodents require further clarity, but the presence of some fossils forms within Sciuromorpha (e.g. *Ischyromys*; Asher et al., 2019) warrant a detailed investigation into the role of the North American continent in the diversification of this crown group. The origins of Myomorpha may also have roots in North America. There are Eocene fossils with ancestral myomorph features in Asia, such as *Pappocricetodon* (Tong, 1997), and also in North America, such as *Heliscomys* (Cope, 1973). The North American fossil *Elymys* in particular, is one of the earliest rodents

to possess myomorph characters and has been used to suggest a North American origin of Myomorpha (Emry, 2007).

Finally, the biogeographic history of Ctenohystrica is of interest as it has the most restricted distribution of the main three rodent groups. Extant species of Ctenohystrica live across South America and Africa, with only Hystricidae and *Laonastes* found in Asia. The diverse Hystricognathi clade (porcupines and caviids) is thought to share a common ancestor with Ctenodactylidae (gundis) and Diatomyidae (*Laonastes*), with the latter two forming Ctenodactylomorphi (Elia et al., 2019). Fossils thought to be closer in appearance to Ctenodactylomorphi than Hystricognathi form part of the 'Ctenodactyloidea radiation', which has fossil roots in the Paleogene of Asia (Antoine et al, 2012). However, the oldest ancestors of Hystricognathi, are found in Africa and South America simultaneously during the Late Middle Eocene (Antoine et al, 2012; Marivaux et al., 2014; Boivin, 2017; Boivin et al., 2019), with Asian hystricognaths not evident until a few million years later in the Late Eocene. If Hystricognathi originated in Asia from a ctenodactyloid-like ancestor, then this would imply large ghost-lineages and therefore gaps in the fossil record.

The influence of the Tibetan Plateau

The role of the environment is another important consideration when trying to understand the diversification of Glires in Asia. Furthermore, it would be impossible to discuss Asian environmental conditions without mentioning the uplift of the Tibetan Plateau. Spanning across central, southern and eastern Asia, the Tibetan Plateau is the largest geological structure of its kind and is elevated more than 4,500m above sea level (Wang et al., 2011). It began to form during the Late Cretaceous as the Indian plate travelled across the ocean from Africa towards Asia. Before impact, the Tethys ocean floor was already experiencing subduction from the approach of the Indian plate, leading to some uplift in Southeast Asia (Royden et al., 2008). The Indian plate then collided with Asia ca. 50 Ma in the Eocene (Torsvik and Cocks, 2016). Subduction of the Indian continental crust under the Asian continent resulted in huge uplift, forming the Tibetan Plateau in the process (Wang et al., 2011). A notable second peak in uplift occurred during the Late Paleogene–Early Neogene, which resulted in the formation of the Tien Shan, Karakorum, Himalaya and Altai mountain ranges (Torsvik and Cocks, 2016). The Tibetan Plateau continues to increase in elevation today (Wang et al., 2011). Severe tectonic activity is thought to be a driver of speciation (Badgley, 2010; Fortelius et al., 2014) and so we might expect peaks in Asian mammal diversity to coincide with periods of significant uplift of the plateau.

Furthermore, the large size and elevation of the plateau is thought to have a considerable effect on weather patterns across Asia. For example, the plateau and surrounding mountain ranges are thought

to drive the seasonal monsoons in Southeast Asia as well as levels of snowfall across Eurasia. There are several case studies which have linked the effects of climate and plateau uplift to the diversification of Cenozoic vertebrates (Wang et al., 2015). These include instances of where cold-adapted species are thought to have diversified in the plateau region during the Pliocene, before expanding into the rest of Asia during the Ice Age (Deng et al, 2011; Deng et al., 2015; Tseng et al., 2013a; Tseng et al., 2013b; Wang et al., 2015). This pattern of adaptation in Tibet and subsequent dispersal when global temperatures cooled has been termed the 'out of Tibet' hypothesis. Although there are many examples of this occurring around the time of the Ice Age, less work has investigated the role of the Tibetan Plateau as a refuge for cold-adapted species further back in time. However, the fossil record of dipodoid rodents points towards the Tibetan Plateau performing a similar role during the Oligocene Glaciations (Pisano et al., 2015).

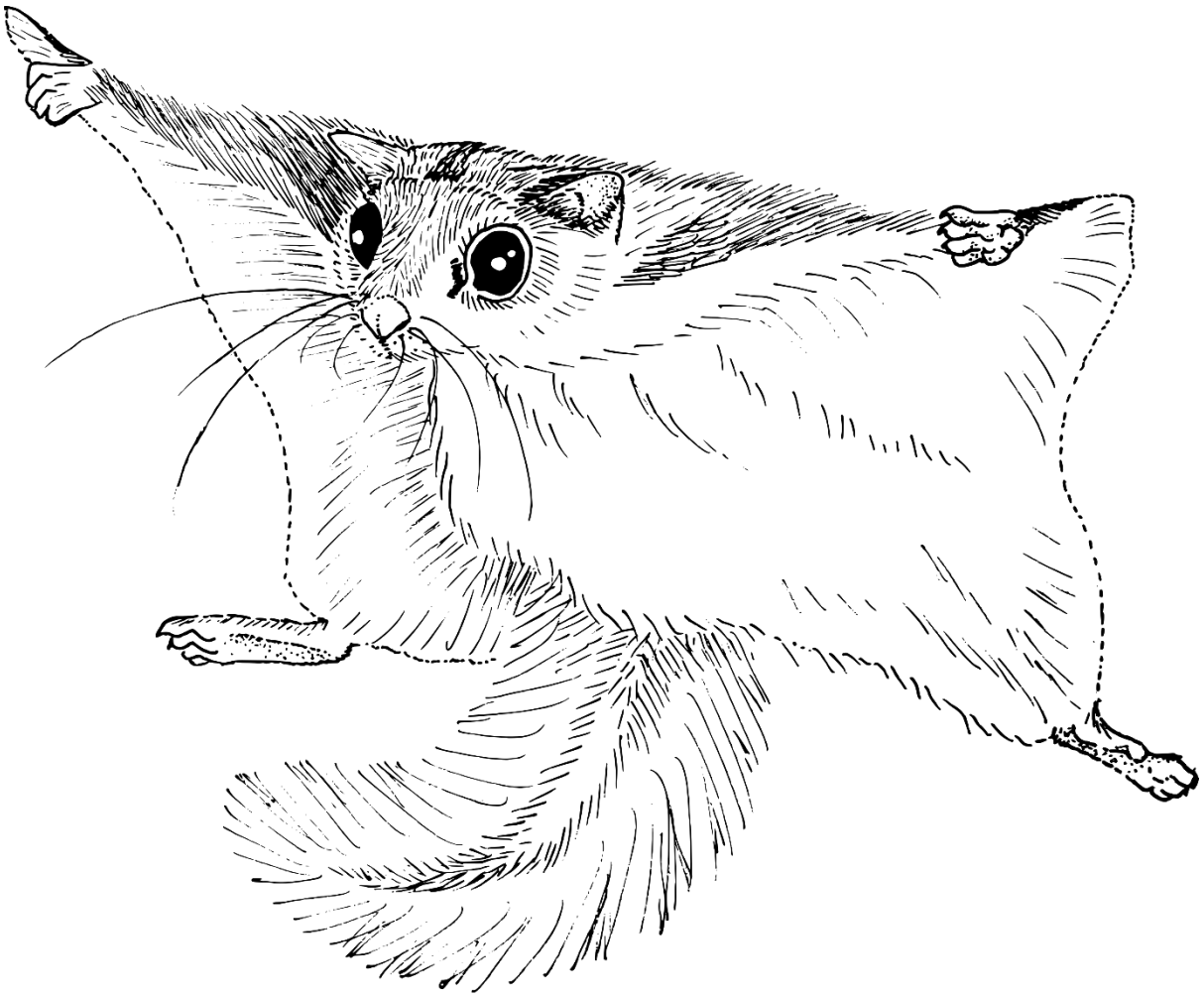
Thesis aims

The aims of my thesis are 1) to examine whether or not the information content of fossil Glires is capable of accurately reconstructing phylogeny, 2) to determine whether morphological partitions in Glires contain different phylogenetic signal, 3) to reconstruct a phylogeny of Glires in order to explore the role of the Asian continent in the origin of the group and 4) to investigate the presence of derived musculature in Eocene fossil Glires.

This thesis encompasses six chapters, including this brief introduction and a short chapter of conclusions. The four middle chapters have been prepared in the form of research articles and so there is some overlap between this summary and their more detailed introductions.

In Chapter Two, I investigate whether or not fossil Glires can accurately reconstruct phylogeny using the artificial extinction method, before teasing apart the nuanced effects of partition biases in Chapter Three. In Chapter Four I use different total-evidence approaches to reconstruct the phylogeny of Glires. I then use these topologies to undertake ancestral range reconstructions to investigate where Glires and groups therein originated. I also combine this information with data on the uplift of the Tibetan Plateau to assess its role in the evolution of Glires in Asia. Finally, in Chapter Five I examine the anatomical characters of 'ischyromyid' specimens to determine whether or not anatomical sciuromorphy was present in Eocene 'protrogomorph' rodents.

Chapter Two: assessing the information content of the Glires fossil record



Glaucomys volans

Abstract

Many lines of research in evolutionary biology rely upon highly degraded fossil data to provide information about extinct organisms. There is some evidence to suggest that the level of missing data in fossils can distort phylogeny, as well as alternative evidence which contradicts this. My study used a heuristic method called ‘artificial extinction’ to degrade the information content of a living species with known phylogenetic affinities to emulate data present in a real fossil. These artificial fossils were then tested to see if they could recover the known phylogenetic relationships of living taxa. I found that artificial fossils based on real fossil Glires (rodents + rabbits) are capable of accurately reconstructing known phylogeny, whereas artificial fossils with randomly coded states (representing noise) do not. Furthermore, the artificial fossils based on real fossil Glires performed just as well as artificial fossils whose morphological partition biases had been removed, suggesting a minimal role of preservational biases in phylogenetic accuracy. However, some fossil taxa that performed better than others with similar amounts of missing data were found to possess less partition bias, indicating that evenly spread data can be more important than the amount of data coded. Finally, I found no evidence for the theory that fossils exhibit 'stemward slippage', further implying that fossil Glires are suitable for phylogenetic reconstruction.

Introduction

Fossils and missing data

The diversity of life we see around us today represents the tip of the iceberg in terms of lifeforms which have lived on the surface of this planet. The majority of species that have ever lived are now extinct, and while it is difficult to know exactly how many species there have been, the fossil record documents an amazing abundance of taxa that are no longer present (Simpson, 1952). The study of evolutionary biology seeks to understand how the world’s diversity came to be and many research questions require information about extinct lifeforms. In nearly all cases, fossils and their geological context are the only tangible evidence of past species' existence.

Fossils provide direct evidence for macroevolutioary processes and so their incorporation into evolutionary studies, phylogenetics in particular, has many advantages. Fossils, in combination with stratigraphy, provide minimum dates for the origin of ancestral clades and so can be used to calibrate molecular clocks and estimate divergence times (dos Reis et al., 2016). A dated tree enables inferences about when living clades first evolved and under what environmental conditions and evolutionary

pressures. Long branches in evolutionary trees, which can be problematic for dating, can also be broken up by including fossil taxa (Giles et al., 2017). Furthermore, the morphological data available through fossils are useful for working out sequences of character evolution and how key anatomical structures might have evolved (Puttick, 2016). Finally, including fossils in evolutionary trees allows us to elucidate the relationships between extinct and extant taxa, and between extant taxa themselves. There are evidently many benefits to incorporating fossil data, but the extent of their usefulness ultimately depends on how accurately they are positioned within the tree of life. Phylogenetic affinities of a given fossil can have very large consequences. For example, two competing hypotheses on the placement of a single fossil might alter the age of a clade by millions of years (Doyle and Donoghue, 1993; Magallón and Sanderson, 2001; Magallón, 2004). This could then have the knock-on effect of changing the sequence of steps by which an anatomical adaptation evolved and how it could be interpreted in terms of biogeography and climate. Deciding how much confidence we have in the phylogenetic placement of a particular taxon is not a problem unique to using fossil evidence, but one which can be scrutinized under the lens of one very prominent fossil problem: missing data.

The process of fossilization inevitably results in data loss as the body of a creature decomposes and is subject to the pressures and conditions under which it was preserved. While this in itself is not necessarily cause for alarm, the observed preservational bias towards hard tissues could be. DNA degrades relatively quickly and currently the oldest molecular data to be extracted from an animal dates to 1.77 Ma (Cappellini et al., 2019). Compare this to the age of one of the oldest vertebrate fossils from the Lower Cambrian (ca. 518 Ma) and the scale of missing molecular data in the fossil record becomes evident (Shu et al., 1999). As well as molecular data, the structural soft tissues themselves are also destroyed very quickly. Truly exceptional preservation of soft tissues such as fur, muscle and protein structures can sometimes be observed, particularly in special sites known as Lagerstätten, but these are the exceptions rather than the rule with most fossils only preserving osteological characters (Allison, 1988). Yet fossils remain our best evidence for past life on Earth.

Handling missing data

How fossils should be incorporated into phylogeny, given their predisposition towards missing data, is something that has been debated for an extended period of time. Since the onset of cladistics and the application of new cladistic methods in the late 1970's, fossils came under a critical eye, most notably from within the palaeontology community itself (Edgecombe, 2010). In his critique on the use of fossils in cladistic inference, palaeontologist Colin Patterson said that the "paucity of characters may severely limit the precision with which relationships may be proposed and tested" (Patterson, 1981, p. 219). Patterson also believed that fossil taxa could not justifiably be interpreted until they

had been assigned to living groups and that the latter should always be the starting point for inferring relationships. This sentiment was echoed by Hennig et al. (1981) and later on by Peter Ax who advocated that "fossils (if any) should be added into appropriate stem lineages after the fact" (1987, p. 229). These views led to fossils being treated as secondary, almost inferior data, which was assumed to have either a negative or little to no influence on evolutionary relationships.

In contrast, Jacques Gauthier, a contemporary of Patterson, Hennig and Ax pointed out that extant taxa themselves are not exempt from missing data (Gauthier et al., 1988). For example, it is easy to imagine a scenario where we might not have a particular gene sequenced for a species or that it could be underrepresented in museum collections and so details of its osteology might be hard to obtain. The difference of course being that while an animal is still extant, the possibility to collect this data remains. To address the effect of fossils and missing data on topology, Gauthier et al. (1988) demonstrated that alternative sister group relationships for reptiles could be achieved by including fossil data. This was one of the first empirical studies of its kind to prove that in spite of their incomplete state, fossils had a measurable influence on evolutionary relationships. This inspired many other researchers to adopt the position that "every effort" should be made to include fossils in analyses from the outset (Donoghue et al., 1989).

Researchers have since attempted further empirical experiments on the effect of missing data, with the focus moving away from proving if a topological effect exists to whether or not it is a desirable one. A series of simulation studies by John Wiens investigated the trade-offs between adding additional morphological characters to a matrix against increasing amounts of missing data (Wiens, 1998; Wiens, 2003a,b; Wiens, 2005). These experiments indicated that large amounts of missing data have surprisingly little effect on tree accuracy as long as the character sample is sufficient. With that said, simulated characters cannot be expected to truly mimic the very complex and often yet unknown processes of evolution and so the need for experiments focusing on real characters is also required. Sansom et al. (2010) recognised this need and so investigated the preservational bias towards soft tissues in living amphioxus and ammocoetes. They performed taphonomic experiments by leaving specimens to decay and assessing the phylogenetic signal of the characters which remained. They found that characters encoding synapomorphies in chordates, according to an *a priori* topology, decayed at a greater rate compared to the less phylogenetically informative characters, an effect known as 'synapomorphic decay bias'. They generalized that this could be the case across all chordates, and with reference to a single cephalochordate fossil, that the fossil record in its entirety is fundamentally biased towards plesiomorphic characters. The effect of this being, they proposed, that fossils occupy erroneously primitive positions within topologies due to stemward-slippage. The

implications of this conclusion for taxa with predominantly soft tissues, such as invertebrates, is of course, much more severe than for vertebrates.

As well as decay experiments, researchers have taken datasets of living taxa with known phylogeny, and artificially reduced the data content by removing characters at random or targeting particular partitions. An example of this would be Springer et al's (2007) "pseudoextinction" experiments. Their dataset consisted of 44 living taxa representing each order of placental mammals. For each taxon they removed all of the molecular and soft tissue characters, whilst keeping all of the osteological characters. This in effect constructs a 'perfect' osteological fossil of an extant taxon. They then substituted whole orders of placental mammals in their dataset with their artificially fossilized counterparts. Their datasets were then subject to phylogenetic analysis before assessing congruence between the resulting topologies and their optimal topology. They found that under a third of placental orders remained in the superordinal groups as defined by the optimal tree. In 2013, Sansom and Wills analysed the effects of missing data across a dataset of 78 vertebrate and invertebrate matrices. They compared the deletion of all soft-tissue characters to the deletion of an identical number of missing characters spread randomly across soft-tissue and hard-tissue partitions. They found that artificial fossils where the soft-tissue characters were excluded performed more poorly in tests of congruence with the optimal tree than artificial fossils where the characters were deleted at random. Furthermore, they found that the movement of taxa within a tree could be predicted based on the type of artificial fossil substitution. The artificial fossils without any soft-tissue characters displayed more stemward slippage than the randomly degraded artificial fossils, which had even levels of crownward and stemward slippage. It is perhaps worth mentioning, that some of their optimal trees were based on small molecular datasets from the 1980s and 90s, and are in conflict with what we know about animal phylogeny today. However, based on their findings, Sansom and Wills (2013) suggested that since real fossils comprise almost of entirely hard-tissue characters, the use of fossils in analyses would lead to misleadingly primitive positions for taxa. They argued that this would lead to spuriously primitive placements of fossils and consequent overestimations of evolutionary rates.

The artificial extinction approach

Taphonomic studies are useful for studying the nature of decay, but living proxies for fossil diagenesis may still struggle to accurately represent the complex conditions and variety of processes under which fossils are formed. Experiments which artificially reduce the information content of living taxa help us to understand the consequences of missing data in phylogenetic analysis, but neither approach makes reference to particular fossil taxa. Another technique to investigate missing data, directly informed by the fossil record is 'artificial extinction' (Asher et al. 2008). This method involves taking a living taxon

(the subject), for which we have a well-corroborated phylogeny, and degrading its data to reflect that found in a real fossil specimen (the template). The artificial fossil then replaces its living taxon counterpart in a new analysis and the resulting topology is compared to the well-corroborated tree. We are in effect asking, if a living species were to be known only from fossilizable data, could we still accurately reconstruct its phylogenetic affinities? This method was first implemented by Asher and Hofreiter (2006) who took a matrix of 10 living tenrecids and made an artificial fossil for each using the fossil *Parageogale* as a template. All characters coded as missing in *Parageogale* were now coded as missing in the artificial fossils. They repeated their tree search with each of the artificial fossil substitutions and found little incongruence with their topology when other living species were left intact. This result boosted the authors' confidence in the placement of *Parageogale*, an African taxon, within the clade of living tenrecs on Madagascar, thus confirming that Malagasy tenrecs are paraphyletic.

Another study, by Pattinson et al. (2015), used a matrix of extant and extinct primates. In this study, three different categories of artificial fossils were made based on real fossil templates. An illustration of how the three types of fossil are made is provided in Figure 3. The first category of artificial fossil they used, called 'real templates', followed the same procedure as Asher and Hofreiter (2006). A living taxon (the 'subject') has any character removed that is coded as missing in a given fossil (the 'template'). The second type of artificial fossil they used were 'random templates'. In this case, the artificial fossil has the same percentage of missing data as the template fossil, but these are distributed randomly across osteological partitions. The final category of artificial fossil used by Pattinson et al. was 'random states'. In these artificial fossils, all characters coded as missing in the fossil template are coded as missing in the subject, but any remaining coded characters are assigned random states. This combination should test if fossils contain genuine phylogenetic signal, as opposed to noise, which is capable of reconstructing known clades according to the well-corroborated tree. Pattinson et al. observed that artificial fossils were more congruent with the well-corroborated tree with increasing amounts of data, although even the most incomplete real template artificial fossils performed relatively well. The random state artificial fossils shared fewer splits with the well-corroborated tree than the real or random template artificial fossils at any level of template completeness. This suggested that primate fossils contain phylogenetic signal capable of accurately reconstructing phylogeny. At high levels of missing data within the fossil templates, the random template artificial fossils displayed better phylogenetic accuracy than the real template artificial fossils. This indicated that fossils which sample across multiple partitions were better at accurately reconstructing phylogeny than if they were biased towards one partition.

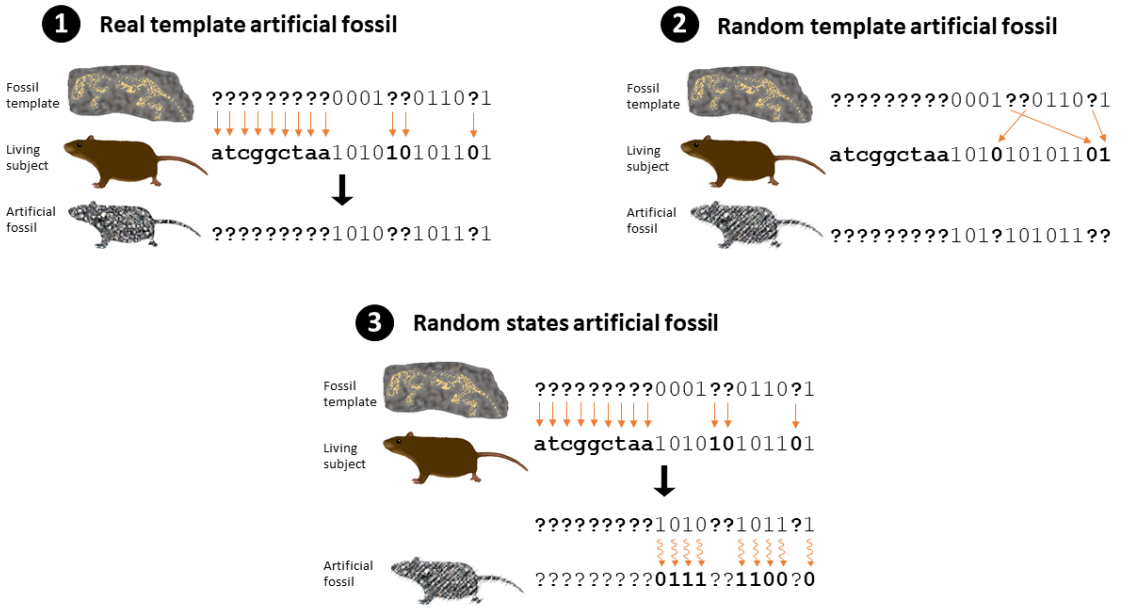


Figure 3: Illustration of how molecular and morphological data are degraded to produce the three categories of artificial fossil used in Pattinson et al. (2015) and this study: 1) real templates, 2) random templates and 3) random states.

Character partition bias

This last finding of Pattinson et al. (2015) is in agreement with later papers written on character partition bias (Mounce et al., 2016; Sansom et al., 2017). The higher density and strength of dental tissues in comparison to other osseous matter often means that teeth are all that remains of a mammal fossil. This observation is likely to be amplified for smaller taxa, which have more fragile remains. As such, there have been many mammal phylogenies based almost entirely on dental characters (Heaton, 1996; Marivaux et al., 2004; Rodrigues et al., 2010; Antoine et al., 2011). However, several authors including Pattinson et al. (2015) and Mounce et al. (2016), have found that sampling across multiple character partitions is beneficial to accurately reconstructing phylogeny. For example, Pattinson et al. (2015) described and implemented a new metric to measure partition bias, called the 'Q metric'. The Q metric is calculated by counting the number of coded characters for each partition and then converting this to a proportion of characters that could have potentially been coded for that partition. The products of these proportions are then summed and divided by the number of partitions. The Q metric measures the extent to which partitioned data are dominated by one element. A score of '0' indicates that the data are present in only one partition, increasing towards '1' as distinct partitions become more complete. They found that fossil templates that were more congruent with the well-corroborated topology than expected for their level of completeness, also had high Q values. It is therefore important when analysing the effects of missing data to also take into account the effects of any character partition biases.

While the artificial extinction method has so far provided compelling results for both arguments relating to the effects of missing data, more can still be done. Sansom et al's (2013) study, although large and taxonomically comprehensive, doesn't make reference to real fossils. Whilst Pattinson et al. (2015) do cover this area, they do so only with one small mammal group and two measures of tree similarity. To address this first point, my study will examine if fossils can accurately reconstruct phylogeny while focusing on the world's largest group of mammals: the rodents and rabbits.

Glires as a study group

Collectively known as Glires, rodents and rabbits comprise almost two thirds of all mammal species, with roughly 2,277 living species described so far (Wilson and Reeder, 2005; Fabre et al., 2015). There are a number of different factors that lend this group to the artificial extinction method and tests of phylogenetic congruence. With the exception of Antarctica, they have colonised every continent and so their fossils can be found in many geographical regions. They also have evolved to occupy a diverse range of niches and lifestyles from arboreal to semiaquatic, cursorial to fossorial. The size range of extant rodents is also disparate with animals ranging from as little as ca. 3 g to as much as 60 kg. All of these factors will affect the conditions and preservational potential of fossil Glires, making them an ideal study group with which to investigate missing data. Furthermore, with advances in phylogenetic methods and molecular data availability, we now have a better picture of Glires systematics than ever before.

Measures of tree similarity

The measure of congruence used to compare tree similarity should be considered thoroughly in artificial extinction experiments as each method measures a different aspect of a tree's structure. The resemblance between any two given topologies is consequently nuanced by the method that is used to calculate it. It is therefore beneficial to examine several measures of similarity. To improve upon the range of similarity indices used in past studies of artificial extinction, my work examines the ability of fossils to reconstruct phylogeny using three different measurements of tree distance: corrected Robinson-Foulds (cRF), shared splits, and quartet divergence (QD). Of these three metrics, the simplest to compute are shared splits and cRF. The shared splits measure counts the number of splits (sometimes called 'bipartitions') present in a tree that are also present in another tree. This method effectively focuses on similarity and ignores unresolved or different relationships. The cRF is a modified version of the Robinson-Foulds metric (RF). The RF method counts the splits that are not shared by two trees, and thus sums the number of splits unique to each tree (Robinson and Foulds, 1981). There are five problems associated with this metric and while they are not unique to this particular measure of congruence, it is undesirable that the RF score is affected by all of them. Firstly,

the RF metric could be considered overly conservative as only perfectly preserved splits are treated as equivalent. Therefore, a very small reshuffle of taxa can disturb many clades simultaneously. It is possible for two topologies with very similar structures to be awarded the worst RF score possible for that topology. For example, two fully pectinate trees can be awarded the maximum RF score (the higher the score, the more dissimilar the topologies) if only the most-nested taxon is grafted and pruned onto the least-nested taxon. This effect can therefore lead to saturation and biased RF scores (Lin et al., 2012). Secondly, the RF metric is somewhat imprecise as the number of unique splits possible is always two less than that of the total number of taxa (Smith, 2019). Thirdly, certain rearrangements of topology can lead to counter-intuitive RF scores. For example, moving one taxon can sometimes give a higher RF score than moving both that taxon and its neighbour to the same position. Fourthly, asymmetric topologies with more even partitions will receive relatively higher scores than more balanced topologies (Smith, 2019). Fifthly, the RF score counts artefactual 'congruence' implied by non-resolved topologies, which is not necessarily desirable. The corrected Robinson-Foulds (cRF) metric implemented in this study aims to reduce some of this effect by scaling the RF score by the resolution of the tree. That is to say, the RF value is normalised against the maximum amount of information that could possibly be resolved on the topology and so punishes poorly resolved trees that are otherwise congruent with better resolved ones. However, even with this adjustment, the cRF is still prone to problems of saturation and does not scale linearly. A different metric that is often proposed, although perhaps less prevalent in the literature due to the ease of calculating RF, is quartet divergence (QD). As the name suggests, rather than breaking the tree down into splits, this metric treats topologies as a series of four taxon statements (Estabrook et al., 1985; Day, 1986; Bandelt and Dress, 1986). The ability to capture all of the tree's structural information means that QD has several advantages over the cRF metric. Quartets contain identical volumes of information, scale consistently as trees diverge and are less likely to reach saturation (Smith, 2019). While metrics such as cRF reward accuracy between trees, the QD metric focuses on information content and precision. As a result, it may be less sensitive to changes in one taxon than shared splits or cRF.

Each metric of tree similarity captures different aspects of tree congruence and so has its own advantages and disadvantages, depending on what features of congruence are being focused on. For this reason, it is good practice to examine several metrics when measuring tree similarity.

Aims and hypotheses

By carrying out artificial extinction experiments and using the various tests of tree similarity mentioned above, my work aims to investigate if fossil Glires are capable of accurately reconstructing

phylogeny. Using several indices of tree similarity should provide a more in-depth understanding of fossil performance than in previous studies, while focusing on Glires provides a large and varied sample with which to test the artificial extinction method. Although it has been argued that fossils distort phylogeny, it is my hypothesis that fossil Glires will contain phylogenetic signal capable of accurately constructing phylogeny, and I will test Sansom and Wills' (2013) claim that fossils are biased towards occupying "erroneously primitive" positions within trees. I will further test if particular categories of morphological characters (such as dental, cranial, and postcranial) exhibit different levels of information content. If a tree resulting from an artificial fossil is more congruent with the well-corroborated tree than expected for the template's percentage of missing data, I anticipate that it will also have a higher Q value than expected. Given that morphological characters are designed to capture a wide variety of phenotypic information (covering both conserved and less-conserved body parts), I expect that in spite of preservational biases, fossil Glires will contain phylogenetic signal capable of recovering crown and stem clades equally.

Methods

Experimental matrix

In order to perform artificial extinction experiments, I assembled a matrix of morphological and molecular characters. The matrix of Asher et al. (2019) contains 219 morphological characters and 14 genes (six mitochondrial, eight nuclear) for 60 living genera and 42 fossils and formed the foundation of my matrix. I added an additional 22 taxa to Asher's matrix (124 in total) based on their completeness and to maximise taxonomic representation. I coded the morphology of one more extant genus (*Cricetulus*) and 21 fossil Glires (*Alagomys, Birbalomys, Branisamys, Dawsonolagus, Elymys, Eocardia, Eogliravus, Eutypomys, Exmus, Franimys, Gaudeamus, Hulgana, Knightomys, Paradelomys, Prolapsus, Protechimys, Sallamys, Sayimys, Tamquammys, Tarnomys* and *Tsinglomys*). I obtained microCT scans of *Cricetulus*, *Franimys, Ischyromys, Paleolagus, Tarnomys* and *Sciuroides* and also examined the microCT scans of *Gomphos*, *Ischryomys* and *Paradelomys* (Rankin et al. 2020). I made virtual reconstructions of the microCT scans using Drishti v.2.6.4 (Limaye, 2012). Other taxa were examined from a mixture of museum specimens and illustrations from the literature. In addition, I added 13 gene sequences for *Cricetulus* to the molecular matrix and aligned them using Opal 2.1.0 and Mesquite (Wheeler and Kececioglu, 2007; Maddison and Maddison, 2019). Appendices 1a and 1c provide information on the specimens I examined. Appendix 1b lists DNA sequences and Appendix 1d provides the combined morphological and molecular data matrix.

Artificial fossils

I followed Pattinson et al. (2015) by making three types of artificial fossil from the data matrix: real templates, random templates and random states (Figure 3). While Pattinson et al. used Python scripts to process the artificial fossils, I adapted R (R Core Team, 2020) scripts from Asher et al. (2019) and Asher and Smith (in prep, 2020). The scripts contain a series of commands to write batch files that can be parsed by the phylogenetics program TNT (Goloboff and Catalano, 2016). A batch file contains the location of the data matrix and instructions for TNT to build several trees. I used the command 'agroup' in the batch file to specify which taxa are extant and which are fossils. I then used the command 'xgroup' to specify the start and end points of the various data partitions, e.g. molecular, crania, etc. I also formatted the batch file to include tree search parameters and modifications to the data matrix. I used the command 'taxcode' to exclude fossil taxa from the tree search. I then used the command 'xread' to define which living taxa to treat as artificial fossils. The R script finds the positions of '?'s in a fossil template and then writes an xread command, which for any given subject taxa, replaces the characters at these positions with a '?'. The R script then writes tree building instructions for each combination of subject and template. I used separate R scripts to make TNT batch files for the real, random template and random state artificial fossils, all of which can be found in Appendix 1e–g. In total, 11,340 artificial fossils were generated, almost twice the amount made possible by Pattinson et al's (2014) dataset.

Phylogenetic analysis

The batch files for TNT created by the R scripts contain specific information on tree search parameters such as the number of replications, type of branch pruning and consensus tree, etc. I used parsimony as the optimality criterion and all characters were treated as unordered and of equal weights. Given the large numbers of trees to be processed, my searches comprised of 100 replications of random addition sequences and held 10 trees after each replication. I performed sporadic tests with more thorough tree search parameters on samples of my data to make sure that the parameters I did use could obtain the same results. Tree bisection reconnection was used for all branch swaps. All most parsimonious trees were kept and a strict consensus tree was calculated for each of the 11,340 tree searches: 60 subjects (not including the root) x 63 templates x 3 types of artificial fossil (real templates, random templates and random states).

Measuring partition bias

To examine if the fossils had biases due to any particular morphological partition, I calculated the Q metric, which was first devised by Pattinson et al. (2015). I divided the morphological characters in my matrix into three partitions based on their common usage in the literature: cranial, postcranial and

jaw-dental characters. The volume of characters in each partition is slightly unequal as 39.7% of the characters in the matrix were jaw-dental, 35.6% were cranial and 24.7% were postcranial. I calculated the Q metric for each fossil template, as well as the percentage of characters coded as 'missing' and the overall percentage of missing characters within each morphological data partition.

Measuring topological incongruence

In order to examine the relative performance of artificial fossils in reconstructing phylogeny, I compared the congruence of artificial extinction topologies to an independent, well-corroborated topology (WCT) of living taxa (Figure 4). I summarised topologies from independent sources. The independent data include; UCEs (Esselstyn et al., 2010; Faircloth et al., 2012; McLean et al., 2019; Swanson et al., 2019), SINEs (Veniaminova et al., 2007; Churakov et al., 2010), microRNAs (Tarver et al., 2016), introns (Rodríguez-Prieto et al., 2014) and RGCs (Mason et al., 2016).

I calculated three types of tree distance metric for each comparison between the WCT and artificial extinction topology: shared splits, corrected Robinson-Foulds (cRF) and quartet divergence (QD). A detailed explanation of each distance metric, including their advantages and disadvantages, is provided in the introduction section above. As TNT produced multiple most parsimonious trees (MPTs), metrics were calculated for each strict consensus of each subject-template combination, before averaging across each template. The R scrips I used to calculate the tree similarity metrics were adapted from Asher et al. (2019) and Asher and Smith (in prep, 2020; see Appendix 1e).

In order to assess whether taxa treated as artificial fossils consistently slip stemwards or crownwards in the tree, I calculated the average node-to-root distance of subject taxa from the root of the tree, before and after artificial extinction using an R script adapted from Asher and Smith (in prep, 2020) and presented in Appendix 1h. As the WCT contains polytomies to reflect the different hypotheses of Glires evolution and does not have representatives for a few of the subject taxa, using this topology for node-to-root calculations would artificially inflate values of slippage for some taxa. To remedy this, I made two additional trees that resolved the polytomies according to competing hypotheses, e.g. the Sciuromorpha-Ctenohystrica clade versus the Myomorpha-Ctenohystrica clade. For 20% of the subject taxa, data independent of the morphology and protein coding DNA in my matrix are too sparse to place them on the WCT with confidence. However, for the purpose of calculating slippage for every possible subject-template combination, I decided to add these 20 taxa and create two completely bifurcating trees (Appendix 1i). I decided their positions within these trees based on long held views of their systematic relationships, as reflected in Wilson and Reeder (2005). I recognise that the placement of 20 taxa based on this data reduces the independence of these two trees and so I will refer to their combined information as the 'reference tree' instead. I calculated the node-to-root

distances for both of these trees and took an average distance for each taxon. I also calculated the average node-to-root distances for each taxon across all MPTs for each combination of subject and template.

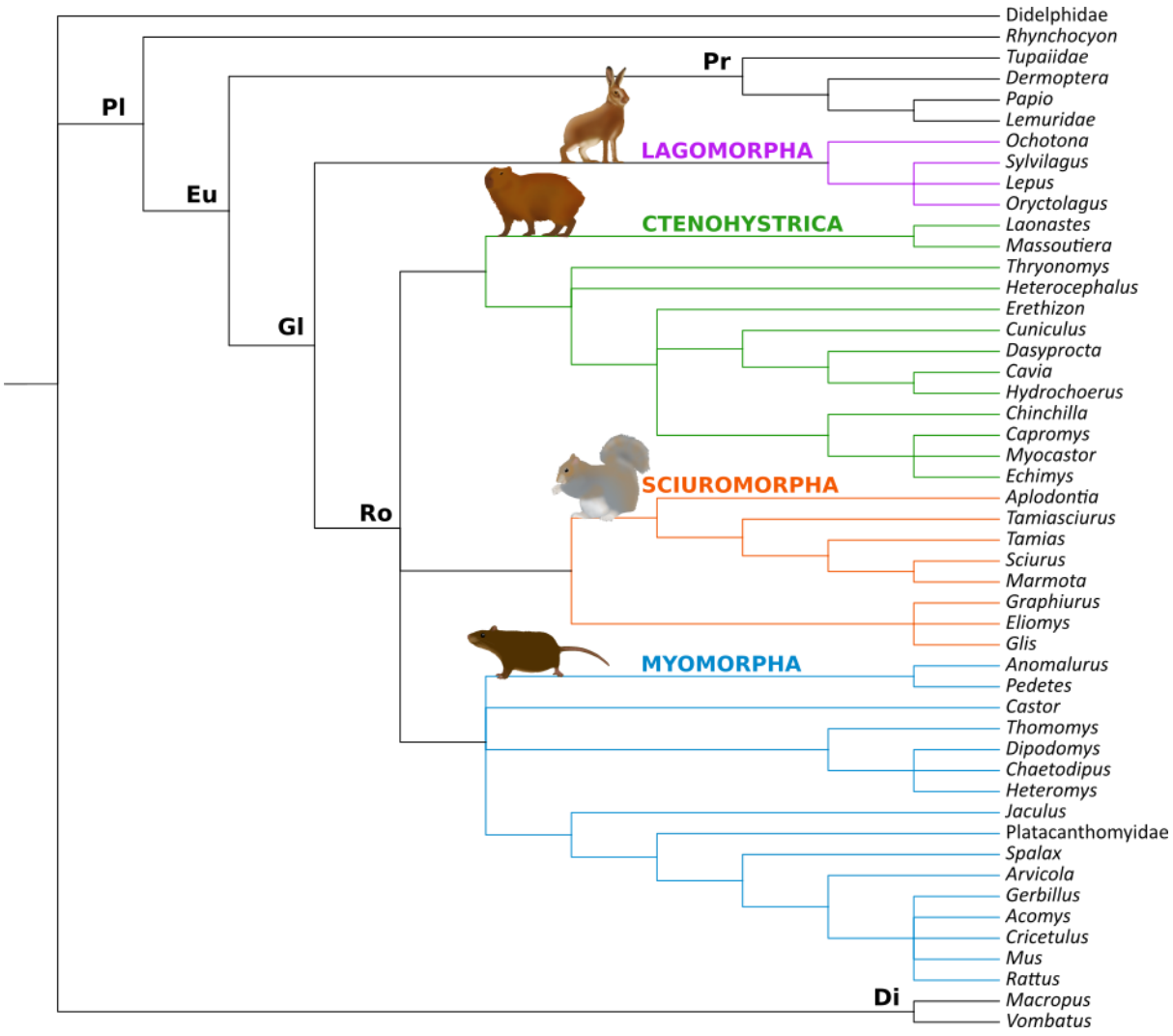


Figure 4: Well-corroborated topology, focusing on Glires and based on a consensus from UCEs (Esselstyn et al., 2010; Faircloth et al., 2012; McLean et al., 2019; Swanson et al., 2019), SINEs (Veniaminova et al., 2007; Churakov et al., 2010), microRNAs (Tarver et al., 2016), introns (Rodríguez-Prieto et al., 2014) and RGCs (Mason et al., 2016). Lagomorpha is shown in pink, Ctenohystrica in green, Sciuromorpha in orange and Myomorpha in blue. Abbreviations are as follows: Di, Diprotodontia; Eu, Euarchontoglires; Gl, Glires; Pl, Placentalia; Pr, Primates; Ro, Rodentia. Branches are of equal length and are arbitrary.

Results

Phylogenetic congruence

Figure 5 shows the congruence between artificial extinction topologies and the well-corroborated tree (WCT) for each type of artificial fossil: real templates, random templates and random states. Results are reported for the three measures of tree similarity used: mean shared splits (Figure 5a), corrected Robinson-Foulds (cRF; Figure 5b) and quartet divergence (QD; Figure 5c). As the molecular and morphological data are only degraded in one subject taxon (the artificial fossil) per artificial extinction topology, we would expect some similarity to the WCT that is not contributed by the artificial fossil's presence. The relative congruence supplied by the artificial fossils is best assessed by looking at the performance of the random state experiments. The random state artificial fossils, unlike real templates or random templates, are coded randomly and so are designed to represent 'noise' rather than meaningful phylogenetic signal. For each measure of tree similarity, the random state artificial fossils recover far fewer WCT splits than either the real or random templates (Figure 5). Additionally, the variation around the mean is much higher for random states, yet they have very minimal overlap with the real and random templates, even at very low levels of fossil template completeness. In fact, even the most incomplete real and random fossil templates have higher congruence values than the most complete random states artificial fossil. This result would suggest, at least for Glires, that real and random fossil templates contain phylogenetic signal capable of reconstructing known phylogeny, whereas random states do not.

Furthermore, the morphological completeness of real templates is positively correlated with mean shared splits ($p<0.001$, $R^2=0.91$), QD ($p<0.001$, $R^2=0.89$) and negatively correlated with cRF ($p<0.001$, $R^2=0.6$). The more complete a real fossil template is, the better its chances of accurately reconstructing the WCT. The morphological completeness of random templates is also positively correlated with mean shared splits ($p<0.001$, $R^2=0.84$) and QD ($p<0.001$, $R^2=0.86$), but negatively correlated with cRF ($p<0.001$, $R^2=0.86$). The variation around the mean congruence values for real and random templates decreases with increasing fossil template completeness. Fossil templates with higher levels of missing data are still capable of recovering the majority of the WCT's structure, but they do so less consistently than more complete templates. This relationship between fossil completeness and phylogenetic accuracy for real and random templates is not linear. Both shared splits and cRF approach an asymptote as more morphological data are added. QD reaches an asymptote around 60% completeness, after which the addition of more morphological information does not increase QD. This could be interpreted to mean that the addition of more morphological characters to fossil templates would further improve shared splits and cRF values, but not the QD of topologies resulting from real

and random templates. Fossils above the QD threshold have between 80–99% of their jaw-dental characters coded, 28–99% of their cranial characters and 2–98% of their postcranial characters coded (Appendix 1j). For this dataset, QD reaches its maximum value once the majority of dental characters are coded. It is worth noting however, that QD behaves differently to shared splits and cRF, as it is less sensitive to changes in one taxon across the strict consensus trees. For example, a single rogue taxon does little to the overall structure of a tree, resulting in a high QD, but depending on where it is placed in the tree, accuracy may lead to poor shared splits and cRF.

Both real and random templates follow similar trajectories as the amount of missing information in a fossil template decreases (Figure 5). The variation around the mean congruence for each fossil template overlaps for real and random templates across all 3 measures of tree similarity. However, there is a notable difference between the congruence of artificial extinction topologies made with real and random templates when the fossil template used has a high level of missing data. When measuring shared splits and cRF, random template artificial fossils are better at reconstructing the WCT than their real template counterparts when morphological completeness is less than 60%. This might be expected given the nature of random templates. The random template artificial fossils are designed to mimic the volume of data in a real fossil, but with information sampled evenly from each morphological partition, which has been found to improve congruence in other studies (Pattinson et al., 2015; Mounce et al., 2016). However, when measuring QD, real templates perform better than random ones with high levels of missing data. According to my QD results, artificial fossils with sampling biases are better at reconstructing the WCT than artificial fossils without sampling biases when there is little information available. As above, this could be due to the QD metric's lowered sensitivity to single taxon changes in the strict consensus trees.

There appears to be no correlation between the morphological completeness of a random state artificial fossil and similarity to the WCT (Figure 5). However, statistical tests report that completeness is negatively correlated to mean shared splits ($p<0.05$, $R^2=0.05$), cRF ($p<0.05$, $R^2=0.1$) and positively correlated with QD ($p<0.05$, $R^2=0.06$). As random state artificial fossils increase in completeness, we are effectively adding more 'noise' to the matrix. This could potentially dilute the phylogenetic signal provided by the living subject taxa and therefore result in the slight decrease in congruence reported by shared splits and cRF. The opposite relationship implied by QD could, as mentioned before, be due to the reduced sensitivity to single taxon changes. For each of these metrics however, the variation in tree similarity that is explained by morphological completeness (represented by R^2) is very low. This reflects the uncertainty caused by large variation around the mean congruence values.

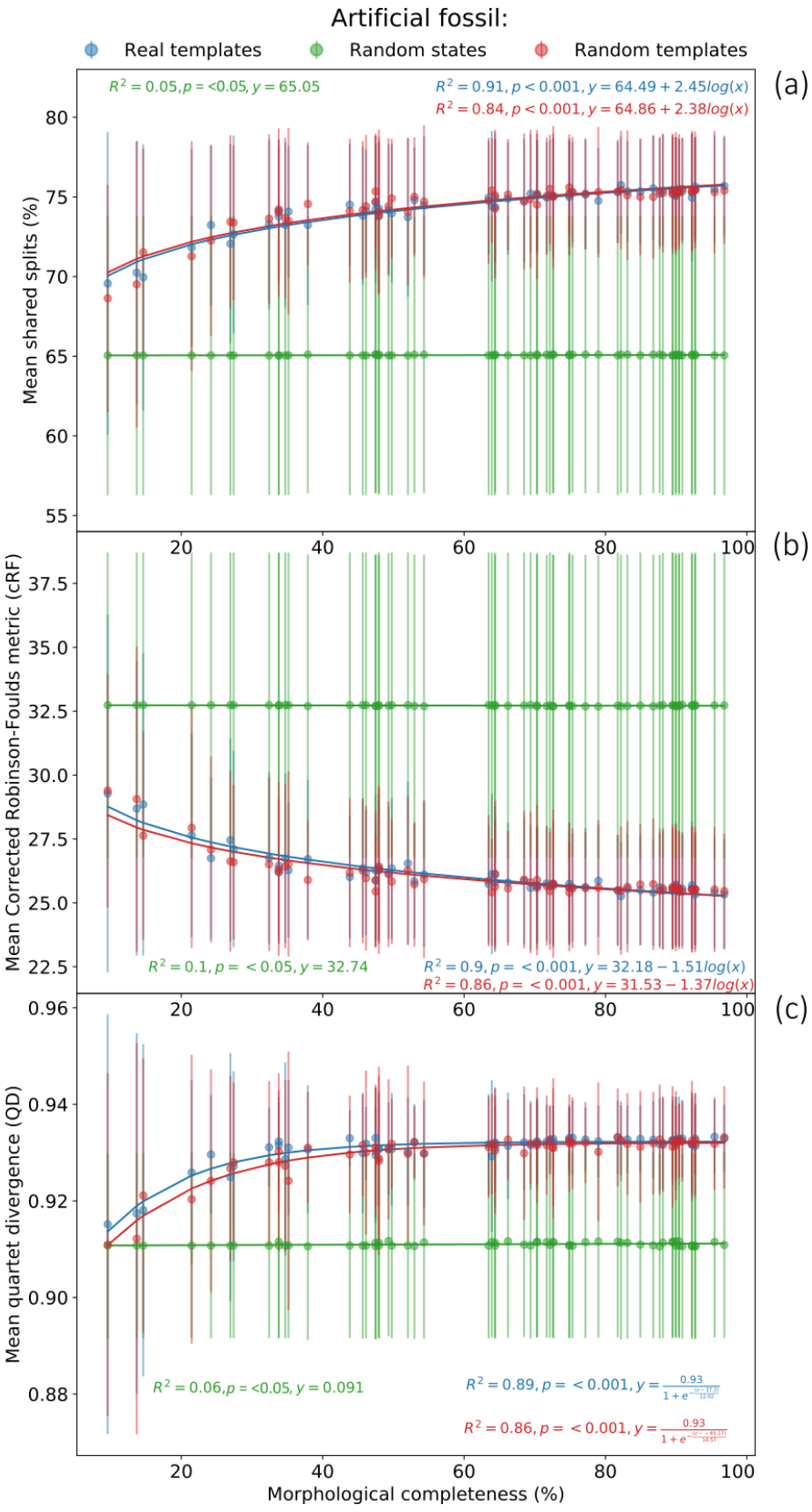


Figure 5: Congruence of artificial extinction topologies to the WCT as described by three measurements of similarity: a) mean shared splits, b) mean corrected Robinson-Foulds and c) mean quartet divergence. Each point represents the mean congruence across all 61 subject taxa and the horizontal bars represent the standard deviation around this mean. The x-axis represents the proportion of morphological characters (out of 219) coded for each fossil template. Real template artificial fossils are coloured blue, random templates are red and random states are green. Lines of best fit accompany each type of artificial fossil along with adjusted R^2 and p-values.

Q metric and partition biases

The Q metric, which measures how data are spread across partitions, increases as fossil template completeness increases (Figure 6). As a fossil template nears 100% completeness, it will eventually sample every character from every partition, thus increasing the Q value. Nevertheless, the results show that a combination of sampling across partitions can increase a fossil's ability to be accurately placed on a phylogeny. We would therefore expect that templates representing data that produce trees with better phylogenetic accuracy than expected for a given amount of missing data would also have higher Q values than expected.

Template taxa, which are considered 'outperformers' in terms of their ability to correctly place living taxa despite fewer coded characters, are highlighted in Figure 6. Different fossil template outperformers can be identified across the three tree similarity metrics, with the most striking outliers found above 50% missing data. *Alagomys* and *Sinomylus* are both outliers in shared splits, cRF and QD measures of congruence. Other fossil outliers below 50% complete include *Protechimys* (shared splits; Figure 6a), *Eogliravus* (cRF and QD; Figure 6 b–c) and *Kulbeckia* (cRF and QD; Figure 6 b–c). Above 50% complete, the near complete heteromyid fossil *Heliscomys* recovers more shared splits than expected (Figure 6a). Interestingly, not all of these outliers have Q values higher than neighbouring templates with similar amounts of missing data. *Alagomys*, *Protechimys* and *Kulbeckia* each have very similar Q values to their neighbours, despite their ability to better place living taxa (Figure 6). *Sinomylus*, *Eogliravus* and *Heliscomys* are the only outliers to have Q values that are higher than fossil templates with similar amounts of missing data. *Kulbeckia* is an interesting case as it recovers an almost identical number of splits in the WCT as the fossil *Eogliravus* (cRF and QD; Figure 6b–c). These two fossil templates have different levels of missing data (52.5% and 56.2% respectively), but almost identical values of Q (0.107 and 0.109 respectively).This could suggest that when missing data are above 50%, even sampling across fewer characters is more important for phylogenetic accuracy than the addition of extra characters at the expense of unbalanced partition contribution. Similarly, if we compare *Sinomylus* to *Protechimys* (Figure 6a), we can see that the latter is more complete than the former, but has a lower Q value and recovers fewer WCT splits. *Protechimys* is more complete than *Sinomylus*, but these are 93% jaw-dental characters, compared to 72% in *Sinomylus*. This is another example of a fossil template with fewer characters than another, which can nonetheless reconstruct phylogeny just as accurately with more evenly spread data. For Glires at least, this would suggest that sampling across multiple partitions is likely to improve phylogenetic accuracy. Having said this, *Alagomys* is consistently identified as an outperformer in my results, yet consists of entirely jaw-dental data. It is not clear why some taxa that are known only from teeth perform better than their peers. It is likely

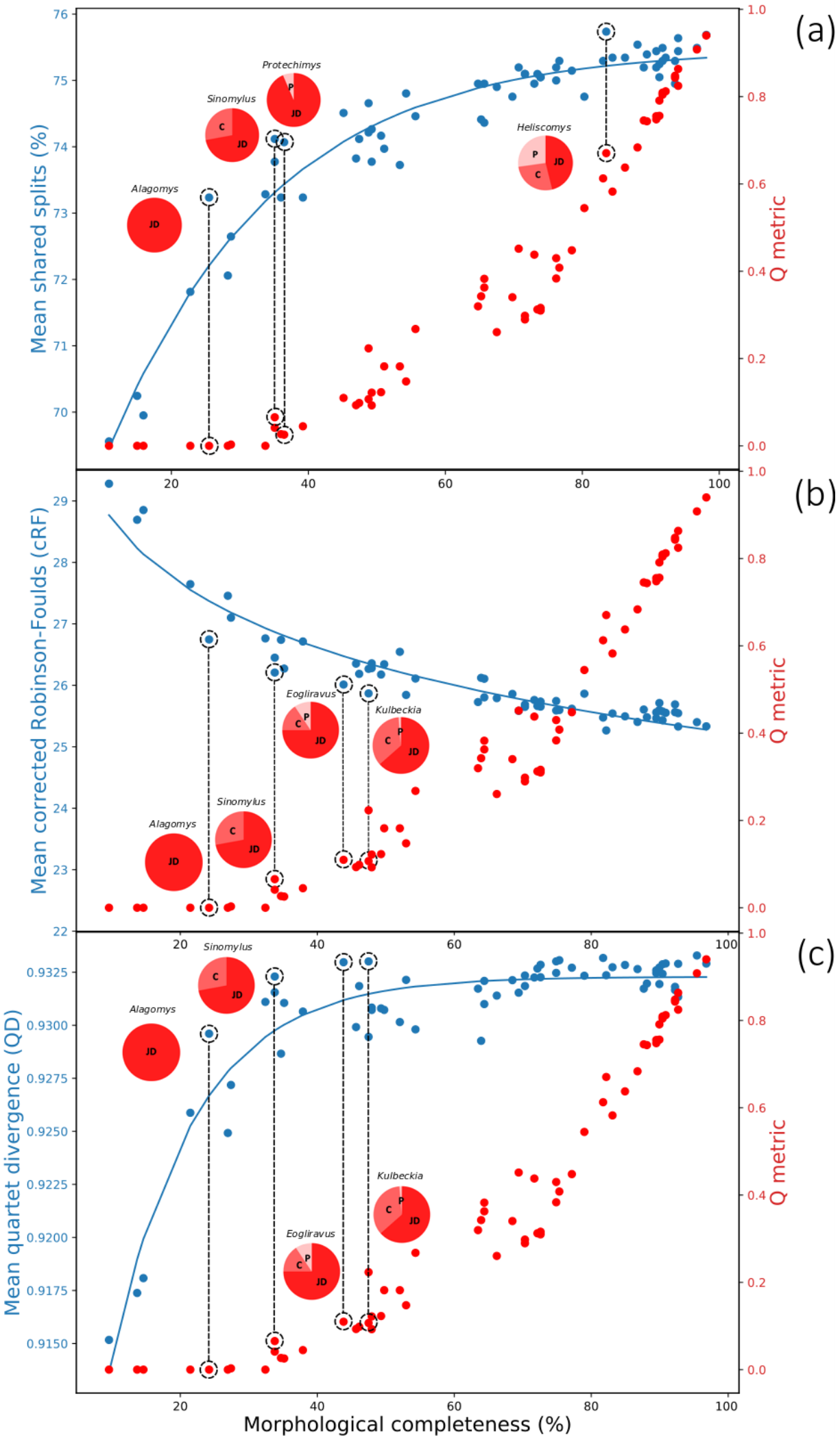
that some dental characters are less prone to homoplasy than others, thus containing more useful phylogenetic signal, causing this disparity among dental only taxa.

Taxon slippage

There is a positive relationship between a fossil template's morphological completeness and the node-to-root distance of its corresponding artificial extinction topology (Figure 7; $p<0.001$, $R^2=0.19$). Artificial fossils made from increasingly complete templates occupy more crownward positions than artificial fossils from less complete templates. It is important to note however, the large range of variation observed, which has led to a very low R^2 score. Figure 8 explores the change in node-to-root distance between living taxa in the WCT and their real template artificial fossils in the artificial extinction topologies. Just over half of the living taxa occupy more crownward positions in the artificial extinction topologies, with one taxon remaining static and the others experiencing stemward slippage. Sansom and Will's (2013) hypothesis that fossils would occupy erroneously primitive positions in the phylogeny is not supported by my data. If my artificial fossils can be taken as proxies for the behaviour of real fossils, then fossil Glires have an almost equal chance of occupying either erroneously derived or erroneously primitive positions within phylogeny.

↓ **Figure 6:** Congruence of artificial extinction topologies to the WCT as described by three measurements of similarity on the first y-axis: a) mean shared splits, b) mean corrected Robinson-Foulds and c) mean quartet divergence. The second y-axis represents the Q metric for each fossil template. The x-axis represents the proportion of morphological characters (out of 219) coded for each fossil template. Four outliers for each measure of congruence have been labelled and are accompanied by a pie chart, which shows what proportion of coded characters are attributed to the jaw-dental (JD), cranial (C) and postcranial (P) morphological partitions. Lines of best fit for each congruence metric (taken from Figure 5) are coloured blue.

Figure 6


(a)
Mean shared splits (%)
Q metric
Protechimys
Sinomylus
Heliscomys
Alagomys
(b)
Mean corrected Robinson-Foulds (cRF)
Eogliravus
Kulbeckia
Sinomylus
Alagomys
(c)
Mean quartet divergence (QD)
Sinomylus
Alagomys
Kulbeckia
Eogliravus
Morphological completeness (%)

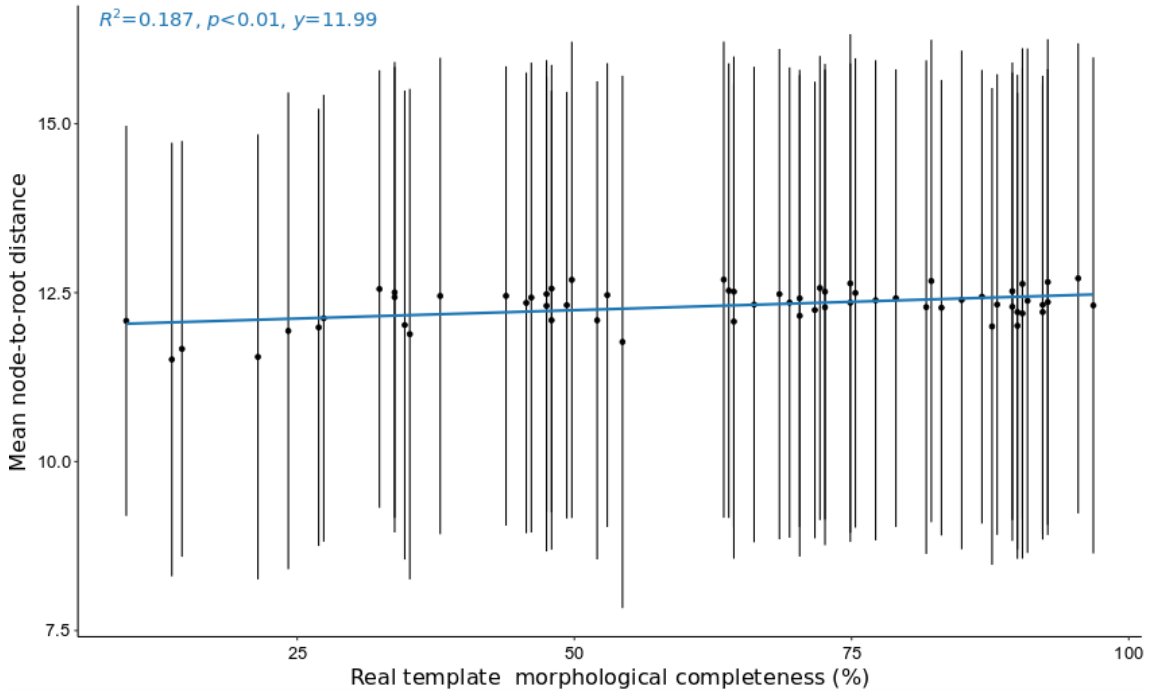


Figure 7: Mean distance from real template artificial fossils to the root of an artificial extinction topology. The y-axis measures the mean node-to-root distance of an artificial fossil to the root of the tree. The vertical bars represent the variation observed around the mean node-to-root distance across the 61 subjects used with each real fossil template. The x-axis represents the proportion of morphological characters (out of 219) coded for each fossil template. A line of best fit is coloured blue along with adjusted R^2 and p-values.

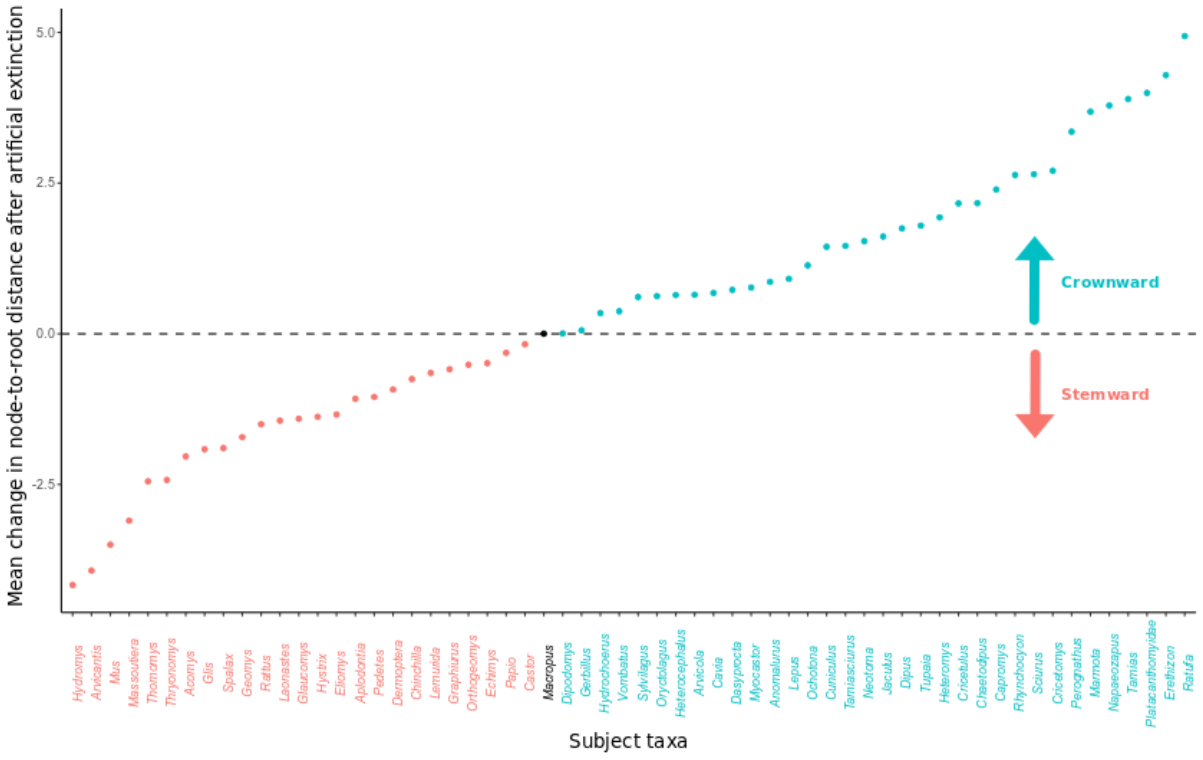


Figure 8: Mean change in node-to-root distance between a subject in the reference tree and each corresponding real template's position in the artificial extinction topologies. Subject taxa whose artificial fossils slip stemwards from their position in the reference tree are coloured red. Those that slip crownwards are coloured blue and those that maintain the same node-to-root distance as in the reference tree are coloured black.

Discussion

The artificial extinction method tests whether fossils retain sufficient signal to accurately reconstruct phylogeny by using artificially degraded living taxa as proxies for fossils in a well-known topology. Whether or not we trust a particular fossil to achieve accurate placement depends on how well it performs in such experiments. In this study, I show that the information contained within even the most incomplete fossil Glires is capable of recovering the majority of relationships within a well-corroborated tree (Figure 5). The most incomplete fossil template in my matrix, the ctenodactyloid *Sayimys*, is coded for only 21 morphological characters (less than 10% complete and exclusively dental characters) and yet recovers 69% of the splits in my well-corroborated tree (WCT), albeit with a large standard deviation. This result is consistent with Pattinson et al's (2015) study of primate fossils, where real templates were also able to recover the majority of splits in their tree of living taxa. My artificial fossils based on random data are not capable of obtaining this level of tree similarity and perform far worse than any of the artificial fossils based on real fossil data (Figure 5). If real fossils were likewise devoid of useful phylogenetic information then we would expect them to behave like these random state artificial fossils, which they do not. This supports the interpretation, at least for Glires, that fossils contain at least some phylogenetic signal. While this may seem like an obvious and perhaps needless deduction, it is important to remember that fossil data has been scrutinized for a long period of time and still continues to be so. As recently as four years ago, Halanych (2016, p. 325) argued strongly that we "should not employ morphology and developmental data to reconstruct the tree" and that instead we should interpret fossils independently of phylogeny. Thirty-five years after the fact, this quote is reminiscent of Hennig (1981) and Ax's (1987) first criticisms on the efficacy of fossil remains. However, artificial extinction experiments support the inclusion of fossil data in analyses of phylogeny.

Nevertheless, preservational biases are still worth investigation. The results of this study indicate that real fossil data are better at reconstructing phylogeny when they are complete (Figure 5; Pattinson et al, 2015). If the remains of extinct taxa contain phylogenetic signal, then any reduction in coded characters is likely to dilute this signal. Conversely, the addition of informative characters is likely to boost it. My artificial extinction experiments show that this assumption only holds up until a certain threshold. When congruence with the WCT is measured by shared splits and corrected Robinson-Foulds (cRF), the incline of the slope gradually decreases as more data are added, and for quartet divergence (QD), this positive incline reaches a plateau. The point of QD saturation occurs when the fossil template is about 60% complete and could be explained by the morphological content of the artificial fossils as well as the behaviour of the QD metric itself. Most of the fossils at this level of completeness already have the majority of their jaw-dental characters coded and so the addition of

extra cranial and postcranial information after this point does not seem to aid phylogenetic congruence. It is also worth noting that QD is less sensitive to individual rogue taxa than shared splits or cRF, as QD focuses on overall structural similarity rather than accuracy. Therefore, with increasing levels of completeness, there may also be fewer instances of multiple tips changing positions, and so less change to QD values than cRF values or shared splits. The differing results between the three tests emphasise the importance of applying more than one tree similarity metric when measuring the congruence of topologies that include fossil taxa. Acknowledging the differences between the various methods, but only reporting the results from one, can allow bias to creep into the results and lead to potentially misleading conclusions about fossil efficacy.

In relation to preservational bias, I hypothesised that fossil templates with higher than expected cRF, shared splits and QD, would also have more evenly sampled characters for their given level of missing data. That is to say, characters sampled across partitions could give fossils an advantage over taxa with similar amounts of missing data from fewer partitions. The reason we might expect this is that any one partition could be affected by divergent selective pressures or noise more than phylogenetic history. This in theory would be overcome when partitions are equally represented (Gatesy and Baker 2005; Thompson et al. 2012). Signal that is common to each partition, but weak when considered alone, can be amplified by including data from multiple sources (Gatesy et al 1999; Gatesy and Arctander, 2000). If we first take into account my random template artificial fossils, which are designed to mimic a fossil with no partition biases, my hypothesis is not very well supported. Across all three metrics of tree similarity, the real and random artificial fossils produce similar results, although at high levels of missing data, there is a small disparity between real fossil templates and the random templates (Figure 5). Shared splits and cRF scores indicate that fossils with evenly spread data slightly outperform real fossils when missing data are extensive (Figure 5a–b). Quartet divergence on the other hand, suggests that real template artificial fossils perform better than the random templates at high levels of missing data, and the disparity is larger too. This might suggest that QD is less sensitive to preservational biases than shared splits or cRF when measuring congruence. The QD metric's robustness towards single taxon movement is less likely to be contributing a factor at high levels of missing data, as more taxa are likely to move positions with increasing missing data. Overall, these results suggest that partition biases account for little disadvantage in fossil Glires. This is an encouraging result for researchers whose mammal groups are particularly affected by an overabundance of certain body parts.

As regards to individual fossil taxa performing beyond what was expected for their amount of missing data, I identified several such taxa from across Europe, Asia and North America (Figure 6). Half of these outperformers had higher Q values and so less partition bias than their peers, whereas the other half

did not. This was the case for outliers identified by each measure of congruence, thus reconciling some of the difference between the various approaches to tree similarity. I also highlighted an example where two outperformers with different completion levels could have the same congruence scores, such as in the zalambdalestid *Kulbeckia* and glirid *Eogliravus* (Figure 6). In this case, *Eogliravus* is less complete, but has a higher Q value than *Kulbeckia*, indicating that that sampling data across partitions outweighs the benefits of being more complete. There are also cases where a more complete fossil template has a poorer congruence score than near neighbours with more missing data. For example, the theridomyid *Protechimys* is more complete than the eurymylid *Sinomylus*, but because it has a relatively higher proportion of jaw-dental data, its congruence scores, despite being relatively high for its level of missing data, are below that of *Sinomylus*. In this sense, fossil templates such as *Sinomylus* and *Eogliravus* could be considered 'overachievers' as they can accurately reconstruct phylogeny, across shared splits, cRF and QD, better or just as well as neighbouring outliers that have more or similar amounts of coded characters.

There is however, one outlier which is composed of only dental characters. The alagomyid fossil *Alagomys*, has relatively high shared splits and cRF for its level of missing data. In addition, *Alagomys* has an almost identical QD score to the ischyromyid fossil *Franimys* (0.9296 and 0.9298 respectively), but is coded for only half of the characters (24% and 54% respectively). This result is unexpected as there are several reasons why we might expect dental only fossils to fare poorly in tests of congruence. In artificial extinction experiments performed by Sansom et al. (2017), dental characters were found to contain different phylogenetic signal to other osteological characters and to molecules. In analyses of morphological partitions, Mounce et al. (2016) found differences in phylogenetic signal between craniodental and postcranial characters, and in Chapter Three of this thesis, I discuss how the jaw-dental characters of Glires have a significantly different phylogenetic signal to cranial characters (Figure 11). In general, it is thought that conflicting homoplasy between morphological partitions can be overshadowed by stronger, shared phylogenetic signal when all data types are combined (Gatesy et al, 1999). At high levels of missing data, it is harder to predict how well a fossil might accurately reconstruct phylogeny and it likely depends on whether the few coded characters happen to share the same signal as the molecular data used to build the WCT. As simulation studies have suggested, high levels of missing data are not necessarily a cause for concern so long as the few coded characters are 'sufficient' for their purpose (Wiens, 1998; Wiens, 2003a,b; Wiens, 2005). In this case, the phylogenetic signal encoded by the dental characters of mostly incomplete fossil Glires does appear to be sufficient for fairly accurate tree construction. My results demonstrate that a paucity of characters should not be used as the sole reason to exclude a fossil from a phylogeny. Often fossils below a certain amount of completeness are filtered out of matrices due to a perceived notion that

they might produce biased and misleading trees. If fossils that have lots of missing data, such as *Alagomys* or *Sinomylus,* are given the chop, then we lose out on the potential benefits that including those fossils would entail. Partiality is introduced to our studies when we first approach taxon selection (e.g. favouring the beautifully preserved *Franimys* specimen over the assortment of loose *Alagomys* teeth), but it is thankfully becoming easier to perform quantitative tests such as artificial extinction to make truly informed decisions. If an independent source of phylogeny is available, I recommend that artificial extinction studies should be carried out *a priori,* in order to determine a fossil's suitability for phylogenetic analysis. It is also my recommendation that future studies which include fossil data should calculate QD as well as the more common measures of tree similarity, in order to thoroughly investigate the value of fossils in topology reconstruction and to make balanced conclusions.

A further bias that could affect our decision on how to treat fossil data is the idea that extinct taxa occupy erroneously primitive positions within a tree, thus misrepresenting phylogeny. As Sansom and Wills (2013) pointed out, such a widespread bias would have the unwanted consequence of inflating evolutionary rates. The view of Sansom and Wills (2013), that missing data within fossil taxa systematically and consistently distorts evolutionary trees through stemward slippage, is not supported by my data. It is true that fossil Glires that are more complete are more likely to occupy crownward positions in the reference tree, but the maximum observed difference in node-to-root distance of my most and least complete fossils amounts to only 1.2 bipartitions (Figure 7). In practice, this is mostly an undetectable difference. In fact, if we look at the mean node-to-root depth of subject taxa before and after artificial extinction treatment, just over half (57%) move crownwards and only 42% move stemward from their known position on the reference tree (Figure 8). If we accept artificially degraded living taxa based on fossil data as proxies for said real fossils, then these results indicate that there is a greater chance of a fossil slipping crownwards than stemwards in a phylogeny. The maximum change in node-to-root distance observed in subject taxa after artificial extinction (4.9 splits to the root; Figure 8) is actually much higher than the difference in node depth between the positions of the most and least complete fossil templates in this study (1.2 splits to the root; Figure 7). This suggests that the living subject taxon used to make an artificial fossil actually has a larger effect on slippage than the influence of the fossil template used.

As a whole, fossil Glires are capable of accurately reconstructing known phylogeny, whether they are preserved well or poorly in the fossil record. Since the particular methods employed here use the information content of real fossils, this provides confidence that these same fossils will obtain the 'correct' position within the tree of life. As with all methods however, there are limitations to the artificial extinction method. My study only substitutes one living taxon for an artificial fossil at a time,

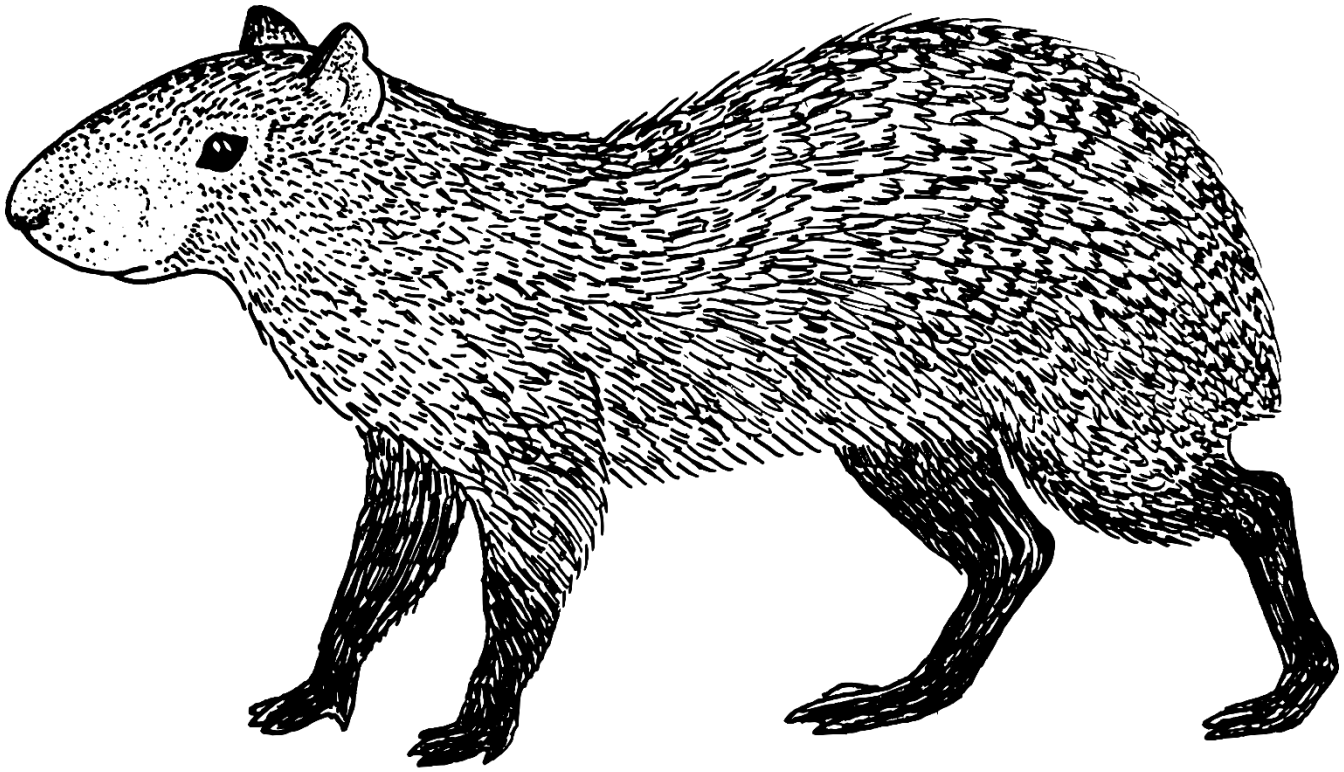
for any given tree search. Studies by Springer et al. (2007) and Brady and Springer (2020), which replace several living taxa at a time, report lower levels of congruence than reported here for single taxon replacement. However, neither Springer et al. (2007) or Brady and Springer (2020) based their artificially fossilized taxa on real fossil templates. Performing artificial extinction experiments on several taxa at a time could mimic more realistic scenarios, as often we wish to incorporate several fossil taxa in phylogenetic analyses. Future research could examine the interaction between missing data and the effects of including multiple artificial fossils in analyses. Another way to enhance the artificial extinction method could be to expand upon the benchmark for phylogenetic 'truth'. The well-corroborated tree used in this study represents congruence across multiple independent molecular datasets and is an improvement on the optimal trees based on single loci used by some studies (e.g. a couple of trees within Sansom and Wills, 2013). For some fossil groups however, a well-corroborated phylogeny may well be possible through ancient DNA recovered from extinct species. Furthermore, an interesting new approach, which compares phylogenetic congruence to the stratigraphic record, has also been developed and could be used for groups without living relatives (Sansom et al., 2018). The artificial extinction method can therefore be extended beyond its current use in this study to further explore the ability of fossils to produce phylogenetic accuracy.

Conclusion

The results of my study offer a positive outlook on the use of fossil data to reconstruct phylogeny. The artificial experiments have shown that fossil Glires can indeed reconstruct phylogeny with acceptable levels of accuracy. Paired with Pattinson et al's (2015) similarly encouraging results for primate fossils, these findings could indicate similar trends across Mammalia. Although preservational biases (towards dental data in particular) have been proposed as a disadvantage of using fossil data, I have not found any serious consequences of partition biases in fossil Glires. However, there is evidence that even sampling across partitions becomes more important for phylogenetic reconstruction than increased volumes of information when the total amount of missing data is high. It is important to note that the different metrics of tree similarity were often in opposition, and so studies that look into the ability of fossils to reconstruct phylogeny should take into account these different metrics. Quartet divergence in particular was less sensitive to preservational biases, which could affect the results of morphological partition experiments or analyses of datasets with large preservational biases. Finally, the idea that missing data in fossil taxa causes them to occupy misleadingly basal positions in phylogeny is not true for fossil Glires. Reassuringly, there is no bias in the movement of fossil Glires towards the stem and very minimal bias towards the crown. The benefits of including fossil data, whether it is sparse or

uneven, far outweighs the costs in terms of benefits for studies of animal evolution. Edgecombe (2009, p. 72) phrased it very well when he concluded that “the exclusion of fossils from phylogenetic analyses is neither theoretically nor empirically defensible”.

Chapter Three: morphological data partitioning and phylogenetic signal in Glires



Dasyprocta punctata

Abstract

Different morphological character partitions (e.g. crania and postcrania) may contain distinctive phylogenetic signal. There have been numerous studies exploring homoplasy and phylogenetic incongruence between partitions, but less work has focused on their impact across individual nodes in a phylogeny. For example, a partition could reflect accurate, historical signal among recently diverged species and genera, but become noisier with deeper nodes and older divergences. In this study, I measure how signal contained within different morphological partitions of Glires (Rodentia + Lagomorpha) varies with node depth across a tree. By measuring the median root-to-unambiguous character change distances across partitions, I found no significant differences between the depths of clades supported by the cranial and postcranial characters. However, when characters are split into smaller groups (e.g. dental, jaw, crania and postcrania), there is a significant difference. For example, dental characters optimize as apomorphies closer to the root of a tree, and rostrum characters towards the tips of a tree. However, knocking out partitions which define basal clades does not generally result in more taxa slipping crownwards than expected. This shows that all morphological partitions of Glires contain at least some signal to support clades across the tree, much of which is likely to be 'hidden signal' when partitions are analysed together.

Introduction

As discussed in Chapter Two, fossils are the only evidence we have for most extinct taxa, but the information that fossils provide is incomplete. Fossils usually represent the osteology of an animal, as soft tissues do not preserve well and additionally, not every bone will withstand the process of fossilization. The fossil record therefore contains large preservational biases. This would not be so troubling if it had been found that all morphological characters carry the same phylogenetic signal. However, increasing amounts of evidence point towards the opposite.

Sansom et al. (2017) found that dental characters are less congruent with trees based on molecular data than osteological characters, with the former being subject to higher levels of homoplasy. Mounce et al. (2016) discovered significant incongruence between the phylogenetic signal of cranial and postcranial characters. In my own results from Chapter Two, I find that fossils which sample partitions more evenly, tend to be more congruent with a well-corroborated tree than fossils which do not (Chapter Two, Figure 6). This result is also seen with morphological data from primates (Pattinson et al., 2015). The cause of homoplasy in morphological characters may concern the functionality of the body parts they encode. Separate anatomical regions can be subject to different

functional selection pressures (e.g., locomotion or feeding), leading to convergence in certain body parts and organ systems across taxa (Ji et al., 1999). In the case of rodent teeth, a strong association between cusp complexity and diet has been hypothesised as the cause for increased levels of homoplasy in dental characters (Evans et al., 2007). Furthermore, we are becoming increasingly aware that anatomical modifications within constrained regions, known as modules, are correlated through time (Klingenberg, 2008; Goswami et al., 2011). Different pressures on modules result in different evolutionary rates across the body and different patterns of homoplasy (Clarke and Middleton, 2008).

A practical implication of this is that if it is not possible to sample characters widely across partitions, results could produce trees that are different to those based on all of the partitions combined (Pattinson et al., 2015; Mounce et al., 2016). Researchers who work with fossil data have no choice but to use the paucity of characters available in the fossil record. However, even when morphological data are bountiful and all partitions are represented, some may still choose to exclude characters that are deemed 'unreliable' on account of homoplasy. The Consistency Index (CI), which measures the homoplasy of a character on a certain topology, can be used to identify highly homoplastic characters, which can then be excluded from analyses or downweighted in some way. An equal weighting scheme gives each character the same weight on topology. Successive Approximations Weighting (SAW) calculates a tree using equal weights and then based on this topology, reweights characters according to their CI, before conducting another tree search with the newly weighted characters (Farris, 1969). This process is repeated iteratively until the weights stabilise. Implied Weighting also gives more influence to characters with low levels of homoplasy, but each additional state change of a homoplasious character is penalised less and less, according to a concave function (Goloboff, 1993). Implementing these different weighting schemes has been demonstrated to have a big effect on tree topology (Wills, 1998; Congreve and Lamsdell, 2016; Lukashevich and Ribeiro, 2020). Common to all of these weighting schemes, is the assumption that a character's likelihood of homoplasy is equal across the whole tree. For example, in equal weighting and SAW, a character change 'costs' as much on lower branches as it does on more superficial branches. In Implied Weighting, although a character's weight is calculated in proportion to how many changes occur, the distance on the tree between these state changes is not taken into account. This however, might not be very desirable and misses out another interesting and understudied aspect of phylogenetic signal.

As pointed out by Goloboff in 1991, it is unrealistic to assume that a character's likelihood of homoplasy is equal across the whole tree, and homoplasy might be more or less of a problem depending on where in the tree it is expressed (Goloboff, 1991; Goloboff et al., 2008; Mounce et al., 2016). Homoplasy at the base of a tree could be more disruptive and result in more incongruence than homoplasy near the tips of a tree. Therefore, it might be good to examine where a character partition

expresses homoplasy before passing judgement on its reliability when reconstructing phylogeny. It also stands to reason that characters in a given partition might optimize as apomorphies differently across a tree, with partitions providing phylogenetic signal in either shallow or more basal parts of a topology. If a particular partition defines clades near the tips of a tree, then absence of this information through fossilization could result in the affected taxa appearing in more stemward positions than they ought to.

In this study, I investigate whether morphological partitions in Glires express phylogenetic signal differently across a tree by examining the distribution of apomorphies. I measure the distance from the root to the node where an apomorphy occurs to determine if partitions provide phylogenetic signal for either more basal or shallower clades. Given the strong functional links of certain morphological characters, I would expect that partitions will differ in where they express phylogenetic signal in the tree. For example, dental data could define clades nearer the tips, as many characters are highly linked to diet, which would not necessarily carry a strong phylogenetic signal through evolutionary time. The mosaic evolution of the rodent masseter system (see Chapter Five of this thesis; Swanson et al., 2019), leads me to believe that characters associated with mastication, such as the zygoma and rostrum, would also be more likely to define superficial relationships. Similarly, I would expect characters less sensitive to changes in ecology and function, such as the middle ear, to optimize as apomorphies closer to the root of a tree. To test how influential these patterns in phylogenetic signal can be for analyses of phylogenetic reconstruction, I examine how the topology of a tree changes when a particular partition is removed. I hypothesise that removing partitions which define clades near the tips, will result in more stemward slippage than removing characters which define more basal clades, whose removal ought to result in more crownward slippage. This information could be useful to consider in analyses where extinct taxa are used, as fossils are often overrepresented by certain partitions, such as dental data.

Methods

Experimental matrix

In order to explore the phylogenetic signal contained within different partitions, I assembled a matrix of morphological characters for living species. While results pertaining to morphological partitions are of particular interest with regard to fossil data, for these specific experiments I focused on extant Glires to eliminate the effect of missing data as a confounding factor. For the matrix I used a dataset of 219 morphological characters for 60 living genera, as published by Asher et al. (2019). I also included

one additional extant genus, *Cricetulus* (Appendices 1a, 2a), sampled with microCT using Drishti v.2.6.4 (Limaye, 2012).

Phylogenetic analysis and character optimization

In order to investigate character optimization, I made a topology based on the morphology matrix. I performed a tree search using the program TNT (Goloboff and Catalano, 2016), carrying out 200 replications of random addition sequences and holding 20 trees after each replication. I chose not to constrain the tree search with a backbone based on molecular data. The goal of this study is to examine the variation in phylogenetic signal provided by morphological data and not to determine if it is capable of accurately reconstructing a well-corroborated tree. I concede that this could result in relationships that are not generally recognised, but constraining the trees in this study would make it impossible to tease apart the phylogenetic signal provided by the morphology and the nuanced contributions of each individual partition when they are removed in subsequent analyses. After the tree search, I calculated a fully bifurcating majority rule consensus tree from 11 fully bifurcating multiple most parsimonious topologies (MPTs). As this consensus tree contains data from all of the morphological partitions, I will refer to it as the 'optimal topology'. This tree can be found in Appendix 2b. I then optimized character changes across the tree from within Mesquite (Wheeler and Kececioglu, 2007; Maddinson and Maddinson, 2019) and used parsimony as the optimality criteria. For each character in the matrix, I counted the number of splits from the base of the tree (excluding the root) to any unambiguous character changes (UCCs). As the state of a character can change multiple times across a topology, I took a median value of the root-to-UCC distances for each character. This effectively measures the depth of the tree where a particular character tends to optimize as apomorphies. Ideally, the depth of character changes would be calculated for each MPT individually and then averaged. However, the calculation of root-to-UCC distances is not yet programmable within TNT and it would be inefficient to calculate by hand given the volume of characters and number of MPTs. I used the majority rule consensus tree as a compromise, as it represents an 'average' tree that characters can be optimized on.

Character partitions and statistical tests

Once character optimization data had been calculated for every character in my matrix, I divided the characters according to different partition schemes. There are many ways in which characters can be split, based either on common usage within the literature or by suspected functional unit. I tested five different partition schemes. My first partition scheme split characters into crania and postcrania, as in Mounce et al. (2016). The second partition scheme split characters into three partitions: jaw-dental, crania and postcrania, as used in Chapter Two of this thesis and by Pattinson et al. (2015). The third partition scheme split the characters into four partitions: dental, jaw, crania and postcrania. The fourth

partition scheme split the characters into 15 partitions: dental, jaw, palate, rostrum, orbitotemporal, fronto-parietal, zygoma, basicranium, occipital, middle ear, axial, pectoral girdle, forelimb, pelvic girdle and hindlimb. The final partitioning scheme split the characters into dental and non-dental characters, as in Sansom et al. (2017).

For each partition scheme, I performed a one-way analysis of variance (ANOVA) to test whether the mean root-to-UCC distances were significantly different between each character partition. I implemented sum contrasts rather than treatment contrasts so that the mean root-to-UCC distance of a character partition would be compared to the mean root-to-UCC distance across all partitions. In other words, this test measures if a character partition's data optimize as apomorphies significantly closer to the tips, or to the roots of a tree, than we would expect given the average root-to-UCC distance across all partitions. I also implemented a Kruskal-Wallis test for each partitioning scheme, which is a non-parametric equivalent of ANOVA. The Kruskal-Wallis test is similar to ANOVA, in the sense that it measures variation among groups of an independent variable given a dependent variable. However, like many parametric tests, ANOVA requires certain conditions to be fulfilled, such as normally distributed variables. The Kruskal-Wallis test does not rely upon distributions and so it is useful to use when the assumptions of ANOVA are not met (e.g. in very skewed data), although it is less powerful than ANOVA (Hecke, 2012). An additional approach to comparing the difference between groups is the Tukey's Honest Significant Difference test (Tukey HSD). While ANOVA and the Kruskal-Wallis test can detect overall significant variation, they do not describe explicitly where those differences can be found. After performing an ANOVA, the pairwise comparisons performed by Tukey HSD will identify which group's means are significantly different from each other (Tukey, 1949). I used Tukey HSD as another method to test if the character partitions contained different phylogenetic signal from each other. This time, the mean root-to-UCC distance for each character partition was directly compared to the mean root-to-UCC distances of other partitions, in a series of pairwise comparisons.

Knockout analyses

In order to test if character partitions cause predictable displacement of taxa when removed (i.e., crownward or stemward slippage), I performed a series of knockout analyses and compared the resulting topologies to the optimal topology. For each partition scheme, I removed one character partition from the matrix and then performed a tree search using TNT, saving all of the MPTs. I repeated this process for each character partition within each partitioning scheme. To see if taxa in the knockout trees had changed position from their placement in the optimal topology, I calculated node-to-root distances for each taxon in the optimal tree and then each of the knockout MPTs. As the knockout TNT tree searches produce multiple MPTs, I took an average of node-to-root distances for

each taxon across the individual MPTs. In order to do these calculations, I used an R script written by Asher and Smith (in prep, 2020), which contains a node-to-root calculation function written by Seraina Klopfstein (Naturhistorisches Museum Baselcan). This script can be found in Appendix 1h. Finally I compared the mean node-to-root distances of taxa between the optimal topology and every knockout topology, and described the movement as either stemward slippage, crownward slippage or no slippage.

Results

My hypothesis that different morphological character partitions in Glires would contain different phylogenetic signal is supported, but depends on how the characters are divided.

Two character partitions

When morphological characters in Glires are split into cranial and postcranial characters (Figure 9), there is no significant difference in median root-to-UCC distance between the two partitions (ANOVA $p>0.05$ and Kruskal-Wallis Test $p>0.05$). Cranial and postcranial characters optimize throughout the tree and not predominantly near the tips or root.

Three character partitions

However, this result changes when characters are split into smaller, more localised partitions. For example, when characters are divided into jaw-dental, cranial and postcranial characters (Figure 10), there is a significant difference in where these partitions optimize on the optimal topology (ANOVA $p<0.05$, Kruskal-Wallis $p<0.01$). A visual inspection reveals that jaw-dental data optimize character changes over a much broader depth of the tree, from root to tips, than the cranial or postcranial partitions. According to the ANOVA results, the jaw-dental partition optimizes as apomorphies closer to the root of the optimal topology than we would expect ($p<0.01$), given the mean root-to-UCC distance across all partitions. Postcranial characters, on the other hand, optimize as apomorphies closer to the tips of the optimal topology than expected ($p<0.05$). The mean root-to-UCC distance of the cranial partition is neither closer to the roots nor to the tips of the optimal topology ($p>0.05$). The Tukey HSD test suggests that the only notable pairwise comparison between these partitions is that of the jaw-dental and cranial characters ($p<0.05$), with the latter exhibiting synapomorphies closer to the crown of tree, and the former closer to the root.

Given that jaw-dental data express character change nearer the base of the tree, we might expect that excluding this data from the matrix would result in more crownward slippage than would be achieved

by removing any of the other partitions. Conversely, we might predict that excluding cranial and postcranial characters would result in more stemward slippage than if we removed the dental character partition. The outcome of the knockout experiments show that it is in fact, more complicated than this (Figure 11). Cranial characters, despite defining clades closer to the tips of the tree than jaw-dental characters, result in more taxa moving towards the tips when excluded (58.3%) than jaw-dental characters (48.3%). However, postcranial characters also define clades closer to the tips of the tree and removing them does result in less crownward slippage (31.7%) than excluding jaw-dental characters. My hypothesis that excluding partitions which provide signal for crown clades would result in stemward slippage is only partially supported.

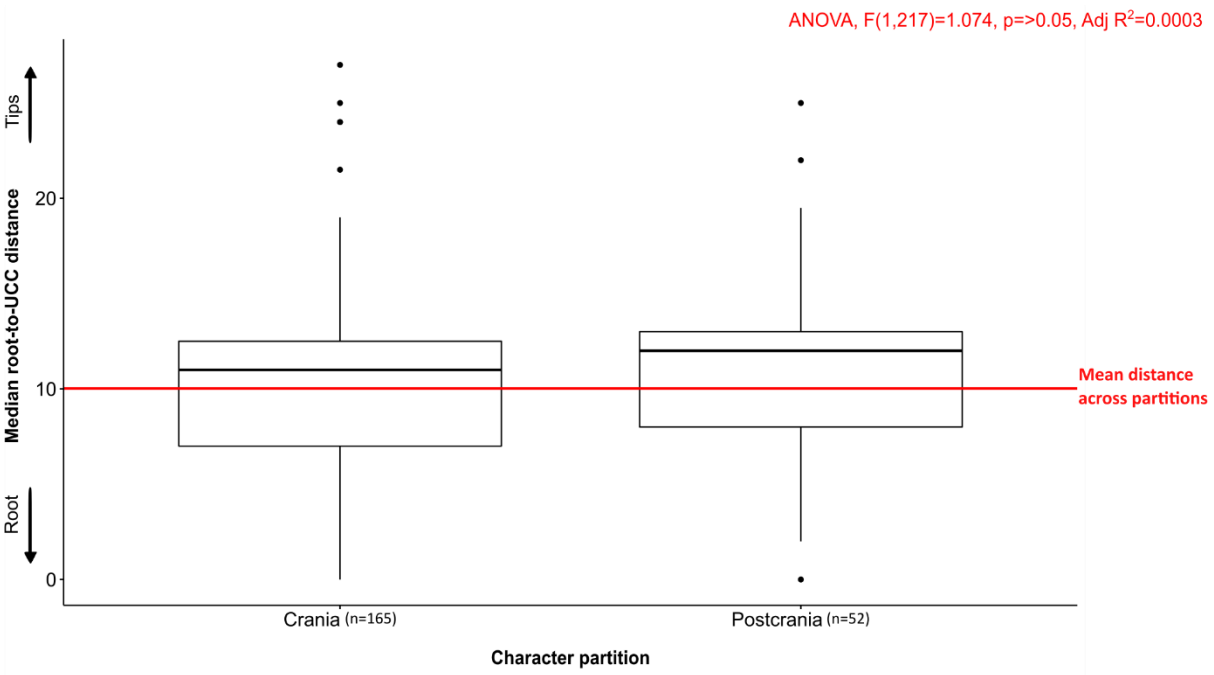


Figure 9: Median root-to-UCC distances of characters optimized on the optimal topology. Characters are split into two partitions: crania and postcrania. The red line represents the mean root-to-UCC distance across both character partitions

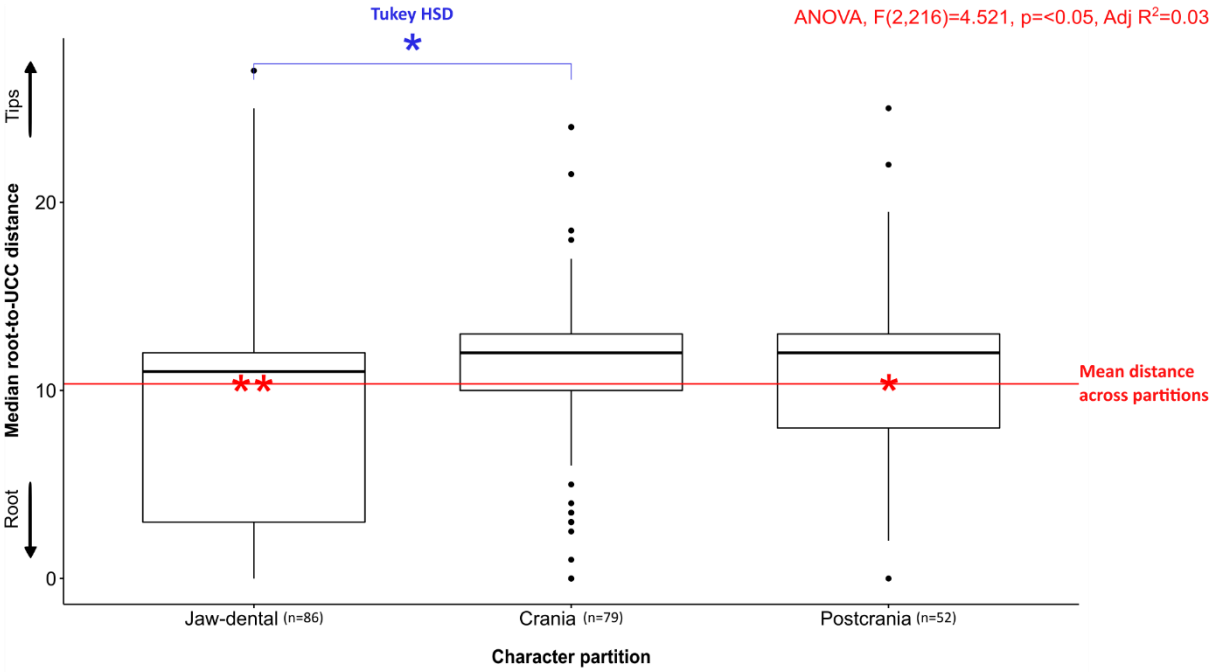


Figure 10: Median root-to-UCC distances of characters optimized on the optimal topology. Characters are split into three partitions: jaw-dental, crania and postcrania. The red line represents the mean root-to-UCC distance across all three character partitions. The blue symbols represent significant Tukey HSD p-values, and the red symbols represent significant one-way ANOVA p-values: * = <0.05, ** = <0.01.

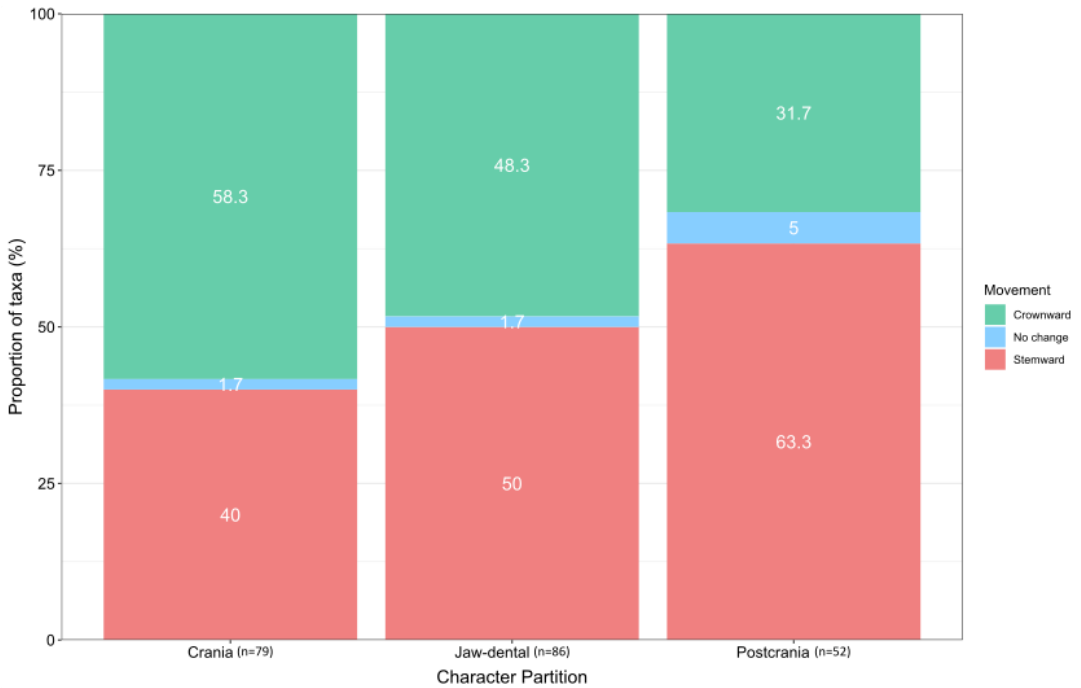


Figure 11: Movement of taxa between the optimal topology and the knockout topology when one character partition is excluded from the analysis. Numbers on the bars indicate the percentage of taxa that have moved positions. Characters are split into three partitions: jaw-dental, crania and postcrania.

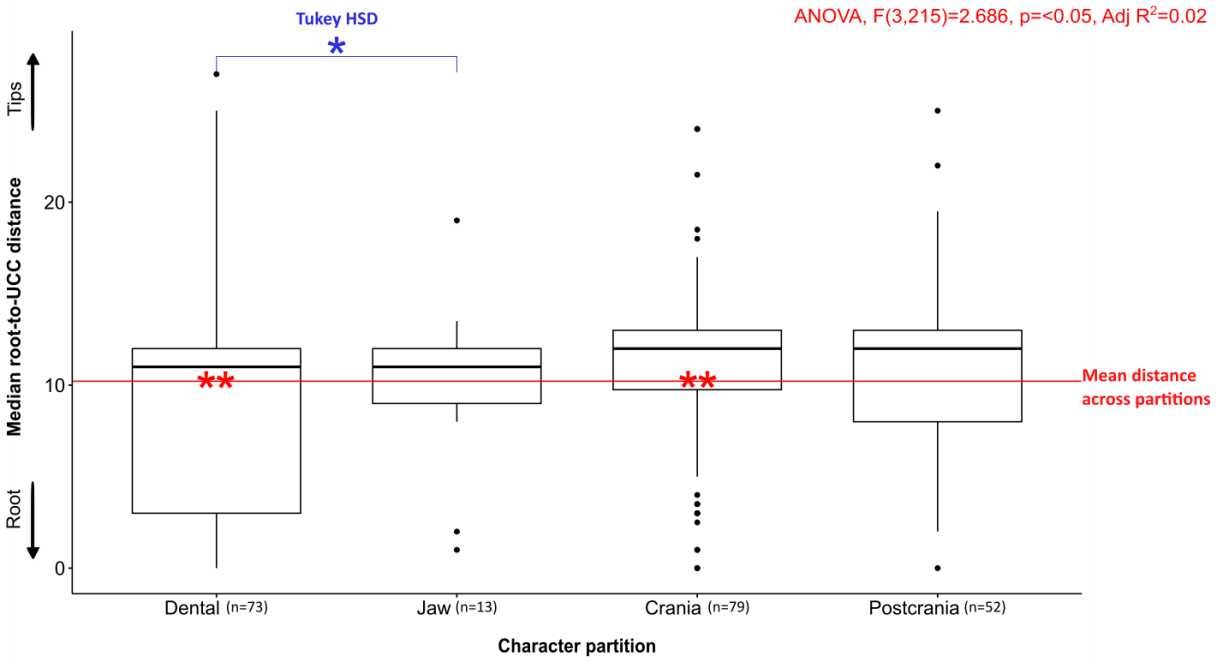


Figure 12: Median root-to-UCC distances of characters optimized on the optimal topology. Characters are split into four partitions: dental, jaw, crania and postcrania. The red line represents the mean root-to-UCC distance across all four character partitions. The blue symbols represent significant Tukey HSD p-values, and the red symbols represent significant one-way ANOVA p-values: * = <0.05, ** = <0.01.

Four character partitions

When characters are split into four partitions (Figure 12), there is also a significant difference in where the characters optimize on the optimal tree (ANOVA $p<0.05$, Kruskal-Wallis $p<0.05$). When dental and jaw characters are differentiated, the Tukey HSD test indicates that their characters optimize at different depths of the tree ($p<0.05$). Dental characters optimize as apomorphies closer to the roots of the optimal topology, and jaw characters optimize closer to the tips. The results of the ANOVA test suggest that dental characters define clades closer to the roots than we would expect, given the overall mean root-to-UCC distance. This contradicts my hypothesis that dental characters would define superficial clades, based on the common selective pressure of feeding. These dental characters are therefore, on balance, conserved. The number of molars, premolars and incisors are a few examples of characters that would be considered highly conserved in Glires, showing relatively little change throughout the clade. The cranial characters optimize closer to the tips of the optimal topology, with jaw and postcranial characters remaining in the middle. In this case, we might expect excluding jaw characters to result in more crownward slippage than removing cranial characters. However, we would also expect that removing dental characters would result in more crownward slippage than removing jaw characters, as dental characters define clades closer to the root. Instead, the opposite of these expectations is true (Figure 13). Excluding cranial characters results in more taxa moving crownward (58.3%) than by removing any of the other partitions. Even the pairwise

comparison highlighted by the Tukey HSD test, between dental and jaw characters, does not result in the slippage patterns we would expect.

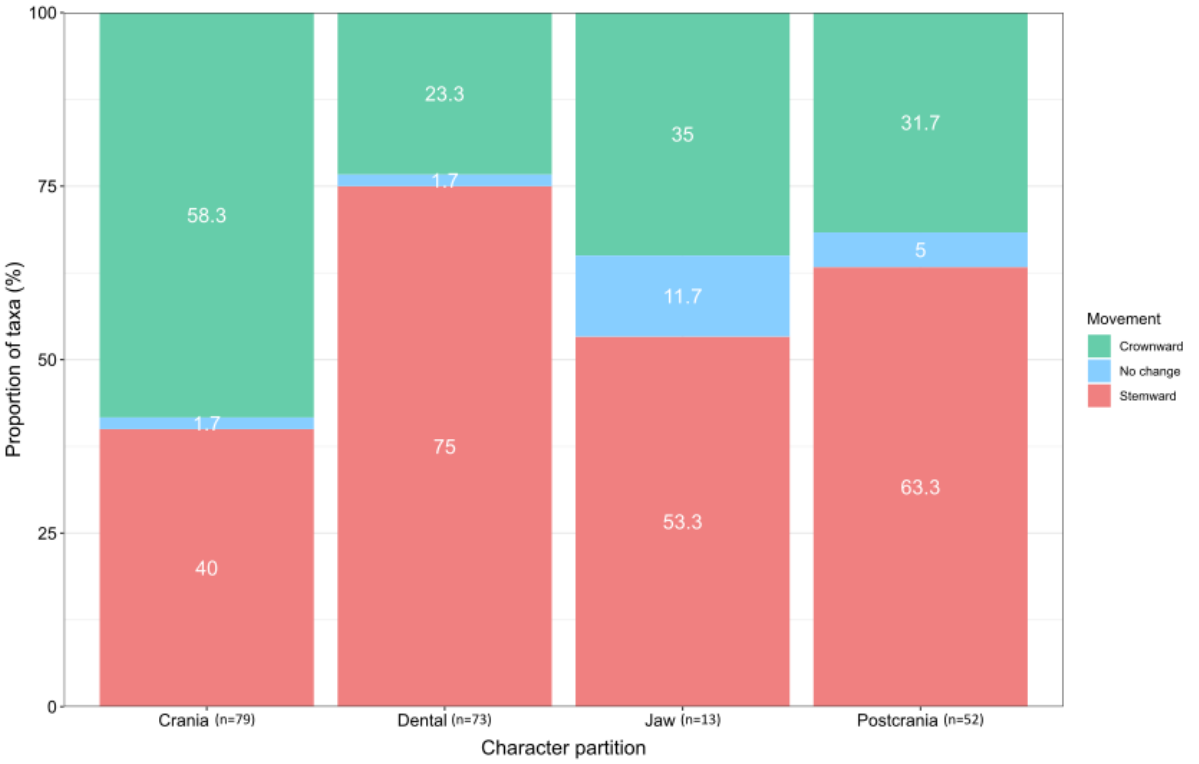


Figure 13: Movement of taxa between the optimal topology and the knockout topology when one character partition is excluded from the analysis. Numbers on the bars indicate the percentage of taxa that have moved positions. Characters are split into four partitions: dental, jaw, crania and postcrania.

15 character partitions

When characters are divided into 15 partitions (Figure 14), there is a significant difference in where characters optimize on the optimal topology according to the ANOVA ($p<0.05$) test, but not according to the non-parametric Kruskal-Wallis test ($p>0.05$). The Tukey HSD analysis finds no significant comparisons between the different partitions, but the ANOVA identifies six partitions that deviate from the overall mean root-to-UCC distance: rostrum ($p<0.01$), orbitotemporal, fronto-parietal, zygoma, forelimb and hindlimb (all $p <0.05$). The results indicate that each of these partitions optimize as apomorphies closer to the tips of the optimal topology than we would expect. For the majority of these partitions, excluding them in the knockout analyses results in the majority of taxa slipping stemward (Figure 15). For example, excluding zygoma characters results in 63.3% of taxa moving towards the root from their position in the optimal topology. However, there are partitions which were not identified as defining crown or basal clades that produce more stemward slippage when excluded, than excluding partitions which were found to define crownward clades. For example, removing dental characters results in 55% of taxa moving stemward, whereas removing fronto-

parietal characters results in only 23.3% of taxa moving down the tree. In general, the percentage of taxa that do not move after a partition is excluded is higher using this partition scheme than in the more condensed partitioning schemes discussed above. This is probably due to each partition containing fewer characters.

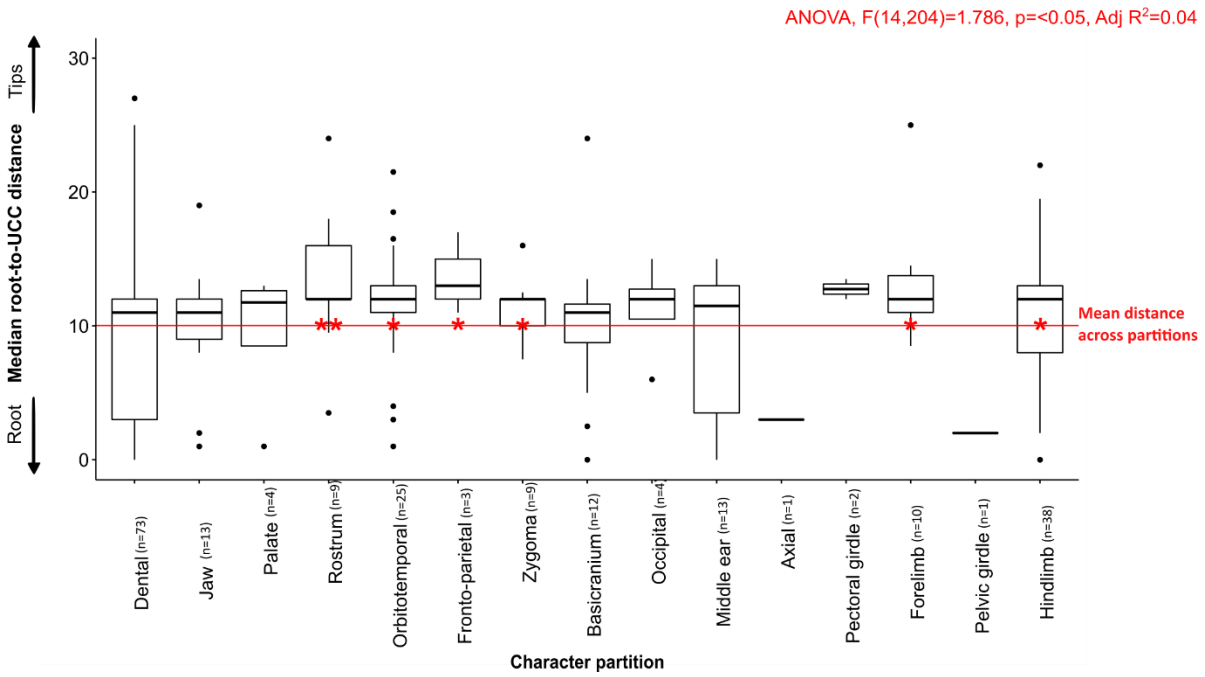


Figure 14: Median root-to-UCC distances of characters optimized on the optimal topology. Characters are split into 15 partitions: dental, jaw, palate, rostrum, orbitotemporal, fronto-parietal, zygoma, basicranium, occipital, middle ear, axial, pectoral girdle, forelimb, pelvic girdle and hindlimb. The red line represents the mean root-to-UCC distance across all 15 character partitions. The red symbols represent significant one-way ANOVA p-values: * = <0.05, ** = <0.01.

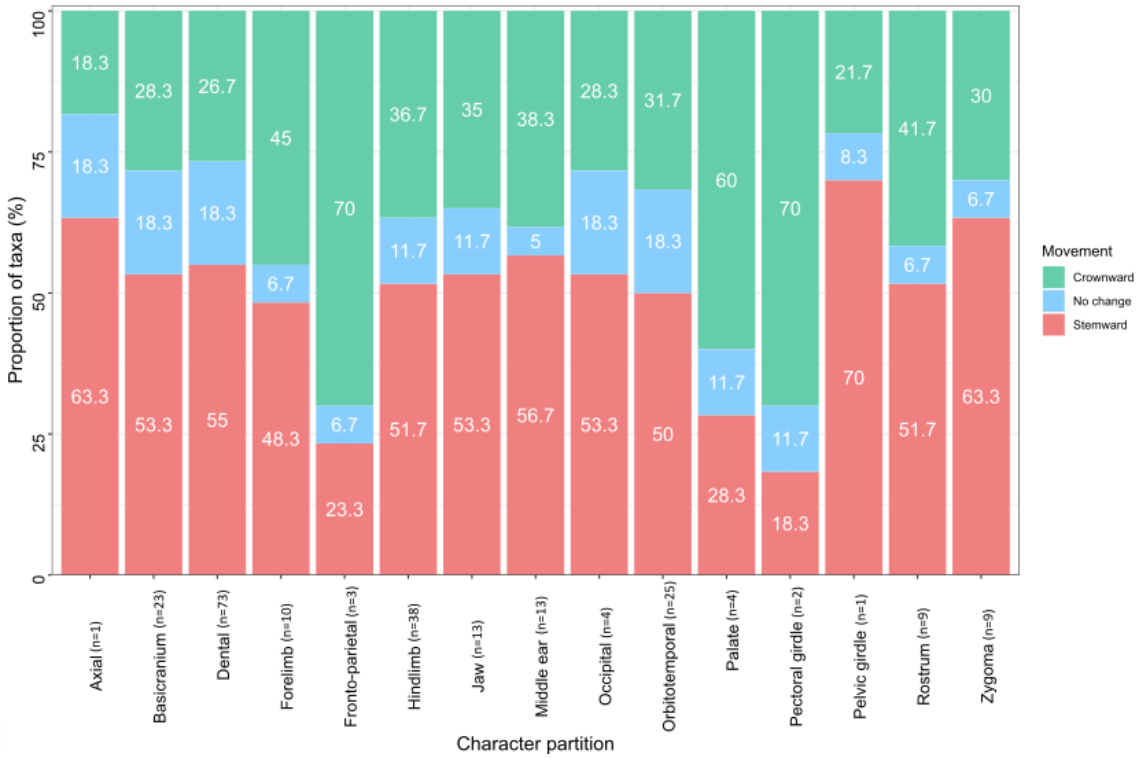


Figure 15: Movement of taxa between the optimal topology and the knockout topology when one character partition is excluded from the analysis. Numbers on the bars indicate the percentage of taxa that have moved positions. Characters are split into 15 partitions: dental, jaw, palate, rostrum, orbitotemporal, fronto-parietal, zygoma, basicranium, occipital, middle ear, axial, pectoral girdle, forelimb, pelvic girdle and hindlimb.

Dental versus non-dental characters

Finally, a large difference in where characters optimize on the optimal topology is noted when characters are split into dental and non-dental characters (Figure 16). In this case, both ANOVA ($p<0.01$) and Kruskal-Wallis ($p<0.01$) tests find a significant difference between the two partitions. Dental characters define clades closer to the root of the optimal topology than expected, and non-dental characters define clades closer to the tips than expected. The Tukey HSD test also confirms that dental characters optimize closer the root of the tree than non-dental characters ($p<0.01$). This result would seem to support Sansom et al's (2017) finding that dental and osteological characters contain different phylogenetic signal. Once again, we would expect that removing the partition that optimizes character changes near the root (dental partition), would result in more crownward slippage than removing the partition that defines clades closer to the tips. This turns out not to be the case here, as knocking out the dental partition results in less taxa moving crownward than knocking out the non-dental characters (Figure 17).

In general, my results show that different morphological partitions exert influence variably across a tree through localised phylogenetic signal near the root or near the tips. My knockout analyses suggest that removing this signal does not produce the results expected. My hypothesis, that removing crown-

defining partitions will result in more stemward slippage than removing partitions which define more basal clades, is only supported when characters are split into jaw-dental, cranial and postcranial characters.

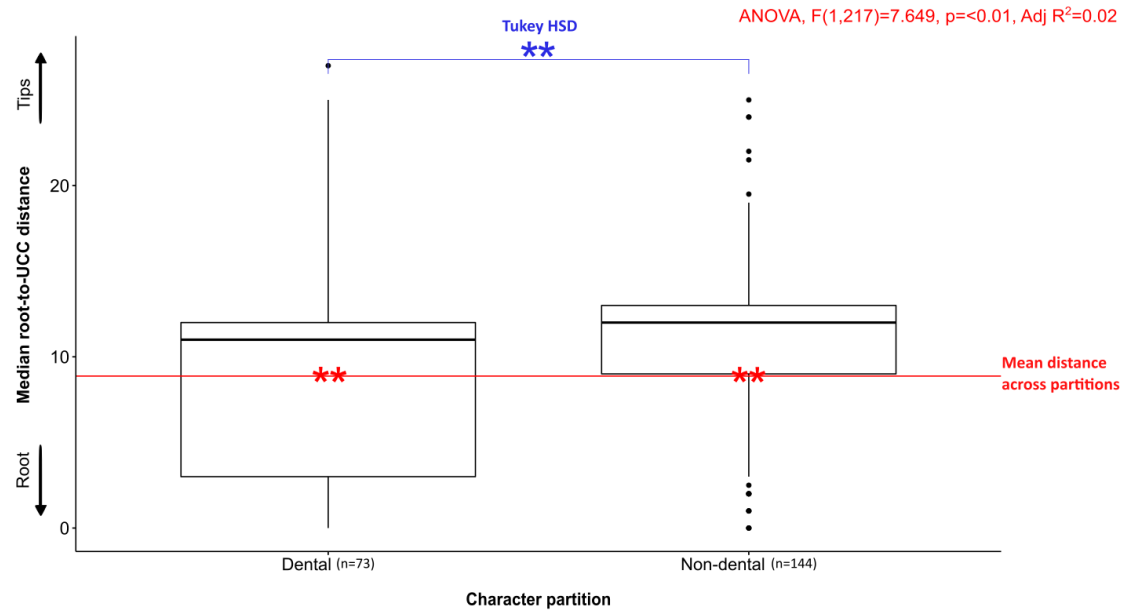


Figure 16: Median root-to-UCC distances of characters optimized on the optimal topology. Characters are split into two partitions: dental and non-dental. The red line represents the mean root-to-UCC distance across both character partitions. The blue symbols represent significant Tukey HSD p-values and the red symbols represent significant one-way ANOVA p-values: ** = <0.01.

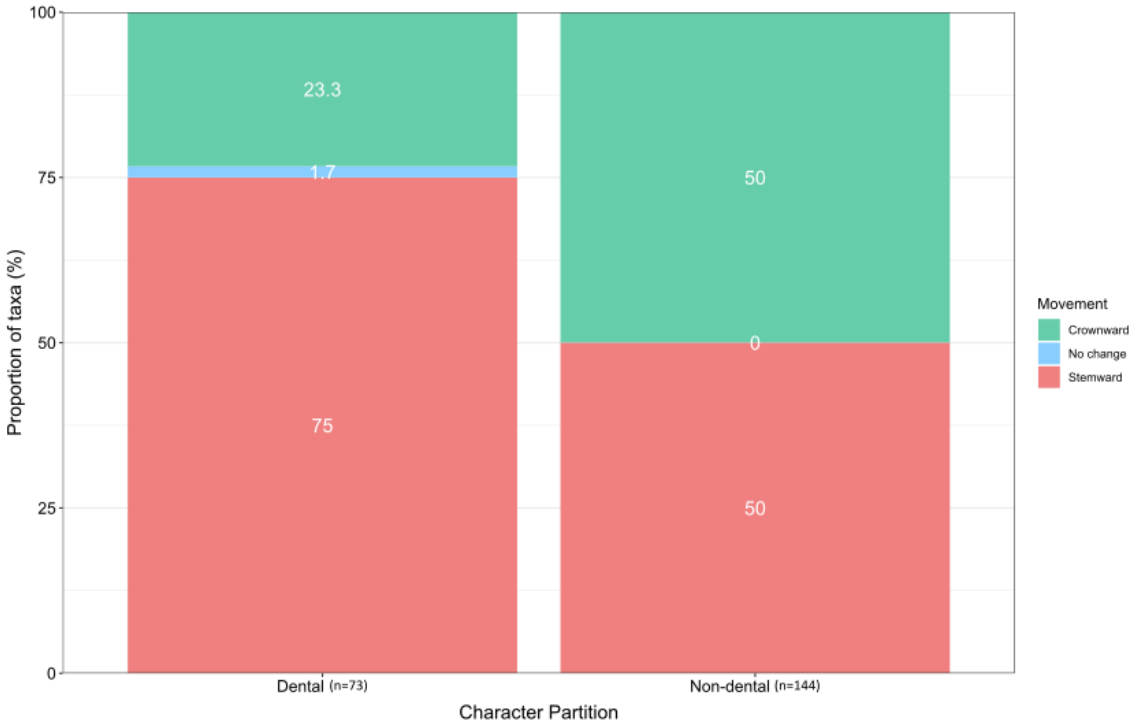


Figure 17: Movement of taxa between the optimal topology and the knockout topology when one character partition is excluded from the analysis. Numbers on the bars indicate the percentage of taxa that have moved positions. Characters are split into two partitions: dental and non-dental.

Discussion

My hypothesis that different morphological character partitions in Glires express phylogenetic signal in different parts of the tree is supported by my data. However, this depends on how the characters are divided. I find no such difference between cranial and postcranial partitions (Figure 9). There is good reason to believe that cranial and postcranial partitions are biomechanically distinct and so we would expect them to be at least partially independent, phylogenetically speaking (Ji et al., 1999). In animals lacking a functional neck region between the head and body (e.g. fish), less incongruence between crania and postcrania partitions have been found (as well as fewer homoplasies), supposedly due to selective pressure from shared locomotion constraints (Klingenberg, 2008; Mounce et al., 2016). In Glires, the subpartitions within my cranial and postcranial partitions likely contain both conserved and less conserved characters, thus providing phylogenetic signal from the roots to the tips of the tree.

As regards to dentition, all of the different partition schemes indicate that dental characters change closer to the root of the tree than those of other partitions (Figures 12 and 16). Sansom et al. (2017), Evans et al. (2007) and Dávalos et al. (2014) all found evidence to indicate that dental data contains different phylogenetic signal to other partitions. My results on the node depth of character changes offer a different type of evidence to suggest this. However, the average node depth of dental phylogenetic signal does not align with what I first anticipated. I hypothesised that dental data would contribute signal to more shallow parts of the tree, but this was not supported by my data. Of the dental characters coded in this study, some are indeed found to define superficial clades at the tips of the tree, such as the presence of P4 lingual cusps in perognathids. However, there are other characters which are very conservative and define basal clades, such as the presence of two incisors per quadrant in Lagomorpha. This finding could reassure researchers who work on mostly dental data that the morphological characters designed to encode dentition provide broad coverage of rodent evolutionary history.

My results also identify partitions which contribute signal to more shallow clades on the tree. Jaw and cranial characters both frequently optimize on more crownward branches than dental characters. (Figure 12). When cranial characters are broken down into multiple partitions, the resolution becomes finer still, with phylogenetic signal from rostrum and zygoma characters clustering towards the tips (Figure 14). Several jaw characters, as well as zygoma and rostrum characters, are directly related to the masseteric muscle system in rodents. As discussed in detail in Chapter Five, the masseteric system in rodents is highly linked to the ability to gnaw, chew and to other dietary specialisations. These adaptations are also widely spread across rodents, with several independent acquisitions and

reversals recorded (Swanson et al., 2019; Rankin et al., 2020). The mosaic nature of rodent masseter evolution could explain the tendency of rostrum, jaw and zygoma characters to provide phylogenetic signal closer to the tips of the tree. There are also conserved partitions within the cranial region, which express phylogenetic signal along more basal branches. The middle ear characters for example, although not significantly different to the mean depth of character change across partitions, do in fact have a large spread of phylogenetic signal across the tree, including close to the root (Figure 14).

Knowing that character partitions in Glires show patterns of localised phylogenetic signal on a tree, we may wish to consider if it is desirable to incorporate this information into analyses of discrete morphological data. In many ways it could be seen as intuitive to weight character changes differently across a partitioned topology, rather than forcing an equal weight for character changes whether they occur near the root or the tips. To not weight characters in this manner is to assume that the presence or absence of phylogenetic signal is equally influential across the whole tree. If a character partition provides phylogenetic signal for basal clades, but is noisier near the tips, we could decrease the cost of character changes along deep branches and increase the cost near tips. In contrast, if a character partition provides reliable phylogenetic signal at the tips of a tree, but is nosier towards the root, we could simply reduce the weight of characters that transform with increasing node depth. The current standards of equal and implied weighting schemes for discrete morphological data have not been greatly expanded upon, except from the implementation of self-weighted optimization by Goloboff in 1997, where character state transformations are weighted depending on congruence. So far, only one weighting scheme has been developed which allows morphological data to be weighted differently across branches. The variable weights criterion was designed to take into account the fact that character reliability might vary in different parts of the tree (Goloboff et al., 2010). Although applying varied weights across an arbitrarily partitioned tree would be possible, this would require prior knowledge of the topology. Instead, Goloboff et al's (2010) method uses the distance between character state changes to calculate the penalty for those changes on the tree being reconstructed. To paraphrase Goloboff (2010), the 'cost' of a transition depends on how many other times this character state transformation occurs in other branches and how far away those branches are from one another. This results in transitions being given more advantage in some parts of the tree than others. The magnitude of this effect is controlled by a topological weighting constant 'q'. When q is large, distant character state changes have as much influence as proximal transformations. In contrast, when q is smaller, character state changes only have an influence if they are very close together. In simulations, the variable weights criterion was found to produce topologies more congruent with a model tree than equal or implied weights, on account of topological weighting (Goloboff et al., 2010).

Perhaps the cost of transformations across a tree could be partitioned to reflect the localised signal within morphological partitions.

It would certainly be interesting to apply a version of the variable weights criterion to future phylogenetic analyses of morphological data for Glires. However, the authors of this method discovered some unexpected and irreconcilable assumptions behind this weighting scheme. They observed that under some circumstances, clades in their trees were being grouped by plesiomorphy as well as underlying synapomorphy (Goloboff et al., 2010). Under simulation conditions where character states evolved independently in several taxa, a tree which grouped them closer together was sometimes preferred over a tree which placed the taxa further apart. Under parsimony, a tree is considered superior to another if it can describe more similarities by common ancestry than its competitor. In the example of the clade being grouped by multiple independent acquisitions of a character, it was not common descent driving this selection. Goloboff et al. (2010) concluded that the variable weights criterion violates parsimony and in fact, any attempt to adjust weights according to distance would have the same consequence. Since common ancestry is the foundation of phylogenetic analysis, it would now seem premature to attempt to weight the morphological characters of Glires according to phylogenetic signal expressed at different tree depths.

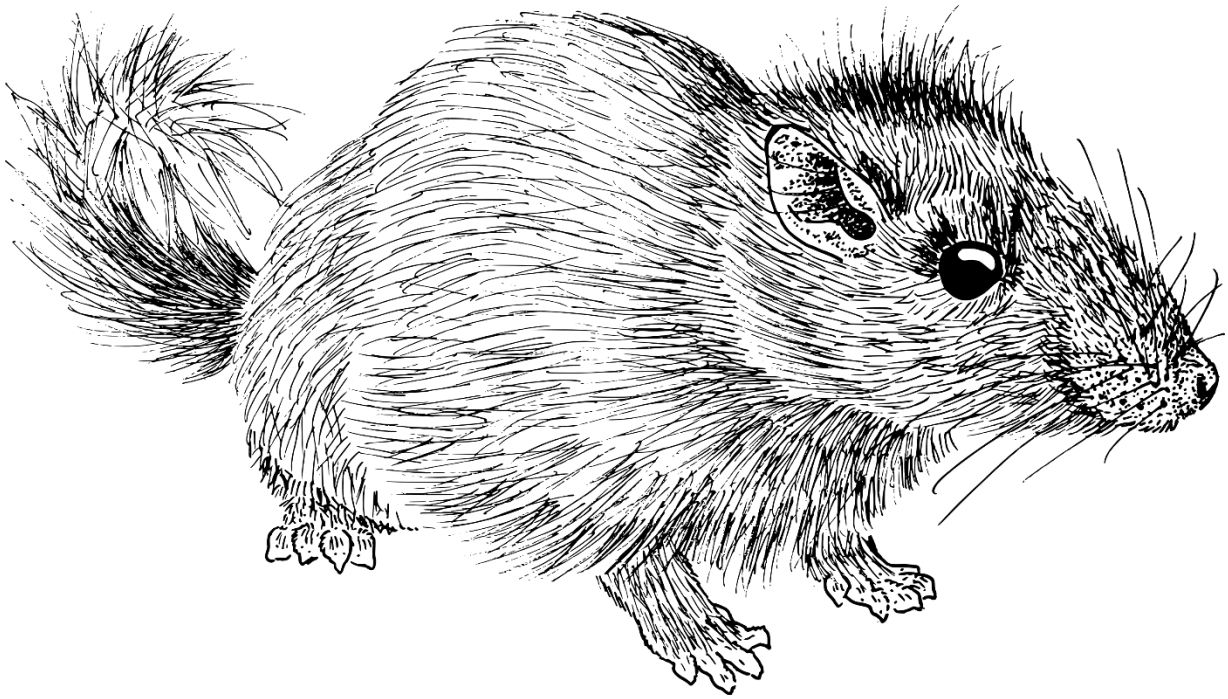
With that being said, the fact that partitions express phylogenetic signal at different tree depths might not warrant the need for complex weighting solutions. My results find little evidence to suggest that the topologically localised phylogenetic signal expressed in partitions has much influence in phylogenetic reconstruction, or at least not in the ways I hypothesised. If a tendency to optimize characters at particular tree depths has a notable effect on phylogenetic reconstruction, we would expect the removal of this localised signal to have consequences for the robustness of clades defined by this data. Data that define shallow clades ought to result in stemward slippage when they are removed. Data that define basal clades should result in crownward slippage when removed. However, my results show that in almost every case, this does not occur. Although movement of taxa in respect to the optimal tree was a consequence of removing any partition in my knockout analyses, overwhelmingly, slippage was in the opposite direction to that expected. The removal of dental characters, for example, resulted in more stemward rather than crownward slippage, despite the fact that they express phylogenetic signal towards the root (Figures 13 and 17). The removal of cranial characters, which define clades nearer the tips, resulted in more crownward slippage than stemward slippage (Figures 11 and 13). The only notable exception to this was when characters were split into jaw-dental, cranial and postcranial partitions. In this case, removing the jaw-dental partition resulted in more crownward slippage than the postcranial characters, whose removal resulted in more stemward slippage (Figure 11). As to why most knockout experiments did not produce the expected

results, it seems likely that while localised phylogenetic signal exists, partitions also contribute signal in other parts of the tree. For dental and middle ear partitions, this is made clear by the large range of their root-to-UCC distances across the tree (Figure 14). When the exclusion of a partition that optimizes nearer the roots results in stemward slippage, it could be that they contain 'hidden signal' for more crownward clades. Hidden signal refers to phylogenetic signal that is weak in any given partition, but when it is combined with the signal from other partitions, becomes strong enough to influence topology (Gatesy et al 1999; Gatesy and Arctander, 2000). In the case of cranial data, which appears to express phylogenetic signal at the tips, we know that it also carries some signal for basal clades (e.g. highly conserved middle ear characters), albeit a smaller amount. However, this smaller, deep-node phylogenetic signal is combined with that of other partitions in combined analyses. When cranial data are removed from an analysis, the small reduction in signal for basal clades might be enough to reduce the combined signal for a particular basal clade, thus resulting in it slipping towards the tips. Hence, exclusion of cranial characters, which appear to define superficial clades, could still result in crownward slippage. Equally, clades near the tips might not move down the tree more than we would expect if the remaining partitions contained enough combined signal to hold their positions. My results could be relevant to fossil data, which are unavoidably missing certain categories of data, such as DNA, soft tissues and are often over-represented by dental characters. Although reduction in phylogenetic signal for some parts of the tree may occur through missing data, it seems likely that even fragmented fossil data will have some phylogenetic signal to cover every part of the tree.

Conclusion

Morphological partitions in Glires contain phylogenetic signal that is localised to different regions of the tree. Partitions such as dental data tend to optimize as apomorphies closer to the root of the tree, whereas cranial characters optimize closer to the tips. Weighting methods could be designed to take this node-depth influence of character partitions into account, but would likely violate the assumptions of parsimony. Knockout analyses however, indicate that removing this localised signal does not force taxa to slip away from these regions of the tree. This suggests that hidden signal is present across partitions in combined analyses of Glires morphological data, which should be beneficial for datasets with under-represented partitions, such as fossil data.

Chapter Four: origins, diversification and biogeography of Glires; the role of the Asian continent and uplift of the Tibetan Plateau



Laonastes aenigmamus

Abstract

The Asian origins of Glires have been suspected for a long time, but the origins of Sciuromorpha, Myomorpha and Ctenohystrica require clarity. Furthermore, the role of the Tibetan Plateau in Asia has been suggested as a driver of evolution in several large mammal groups, but little work has investigated its influence in deep-time or for smaller-bodied mammals. In this chapter I used total-evidence phylogenetic methods to analyse a combined matrix of fossil data and protein coding loci, as well as a matrix which includes ultraconserved elements (UCE). I then performed analyses of ancestral range estimation on the resulting topologies to trace the geographic history of Glires and combined this data with known periods of intense uplift in the Tibetan Plateau region. My results corroborate an Asian origin for Glires, but are unable to resolve the root of Rodentia. The positions of several key fossil taxa differ between the topologies and have resulted in different hypotheses of biogeography for the main three rodent groups: pitching either a North American origin for Myomorpha and Sciuromorpha, or almost anywhere other than North America. Both topologies however, confirm an important role of the Tibetan Plateau in the evolutionary history of Glires, which possibly acted as refuge for cold-adapted taxa, which later diversified during the Oligocene and Miocene glaciations.

Introduction

Lagomorphs and Rodents (collectively known as Glires), are the most speciose group of living mammals known today and are represented by around 2,500 species (Wilson et al., 2016). Lagomorpha contains the rabbits, hares and pikas and is the smaller of the two extant groups (around 95 species). The rest of the species belong to Rodentia, which comprises of three living groups: Sciuromorpha (squirrels, dormice and the mountain beaver), Ctenohystrica (gundis, porcupines, blesmols, degus, agoutis, spiny rats, chinchillas and guinea-pigs) and Myomorpha (anomalurids, beavers, kangaroo rats, mole-rats, hamsters, voles, and true rats and mice). Glires are an incredibly successful group, not just in terms of numbers of species. They are found on every landmass except Antarctica and cover most locomotor adaptations and diets, thus filling many niches.

The timing and geographical setting of the origin of Glires is an important topic, but there has been debate over where and when Glires first appeared. Most molecular clock analyses suggest either a Late Cretaceous (Hallstrom and Janke, 2010; dos Reis et al., 2012; Tarver et al., 2016) or an early-Cenozoic origin (Douzery et al., 2003; Kitazoe et al., 2007; Wu et al., 2012). Of course, knowing the membership and fossil taxa of a particular group is key to calculating the date of its origin. Fossil

occurrences provide minimum dates and so the inclusion of a fossil taxon could in theory move the beginnings of a clade backwards or forwards respectively, by many millions of years. The locations of fossil finds are also vital to understanding the geographic history of a group, as the present-day distribution of living species may not reflect past biogeography. In order to truly understand the evolutionary history of Glires, we need to examine fossil data carefully and consult phylogenies based on information from both living and extinct taxa.

Origins in Asia and the oldest Glires

The oldest mammals that are recognisably Glires date back to the Early Paleocene, just after the K-Pg extinction event (Table 1). From this time until the Late Paleocene, the remains of these ancient animals are known only from a single Chinese region, Qianshan in Anhui (Li, 1977; Li et al., 2016). Most of these basal Glires are attributed to one of two groups, eurymylids and mimotonids (sometimes referred together as 'Mixodontia'), which are thought to possess ancestral Glires characters. Traditionally, mimotonids have been defined by two pairs of ever-growing upper and lower incisors (as in lagomorphs), conforming to the morphological grade 'Duplicidentata' (Sych, 1971). The oldest mimotonid is *Mimotona* from the Early Paleocene (Li, 1977; Dashzeveg and Russell, 1988), followed by *Mina* in the Middle Paleocene (Li et al, 2016) and later on, *Amar* in the Late Paleocene (Dashzeveg and Russell, 1988). Eurymylids on the other hand, have been defined by a single pair of ever-growing upper and lower incisors (as in rodents), conforming to the morphological grade 'Simplicidentata' (Sych, 1971). The eurymylids are sometimes referred to as 'non-rodent simplicidentates' (Dashzeveg and Russell, 1988). The oldest eurymylid is *Heomys* from the Early Paleocene (Li, 1977; Wang et al., 2016), with many other taxa appearing in the Late Paleocene from the more northerly Mongolian Plateau: *Eomylus* (Meng et al., 2005), *Khaychina* (Dashzeveg and Russell, 1988), *Sinomylus* (McKenna and Meng, 2001) and *Taizimylus* (Mao et al., 2017), to name a few. However, this version of classification is oversimplified. There are interesting cases where fossils thought to be typically 'eurymylid' in appearance can have duplicidentate-like incisors, such as in *Sinomylus* (Lopatin and Kondrashov, 2003). It is also not certain if eurymylids or Mixodontia as a group are monophyletic. In phylogenetic analyses, eurymylids are often positioned on the rodent stem and mimotonids on the lagomorph stem (Meng and Wyss, 2001; Meng et al., 2003; Asher et al., 2005).

Possible North American origins

It is not until the Late Paleocene that we start to see fossil Glires outside of Asia (Table 1). The wave of new species that proceed the eurymylids belong to two groups: the 'Rodentiaformes' and 'Ischyromyidae'. Rodentiaformes resemble eurymylids and mimotonids as well as crown rodents (Fostowicz-Frelik, 2020). Three genera of Rodentiaformes are currently recognised and belong to the Late Paleocene family Alagomyidae: *Tribosphenomys* and *Neimengomys* from Inner Mongolia (Meng

Species	Age							References and comment
	Early Paleocene	Mid Paleocene	Late Paleocene	Early Eocene	Middle Eocene	Late Eocene	Oligocene	
Basal Glires								
Amar aleanor			■					Dashzeveg and Russell (1988)
Eomylus bayanulanensis			■					Meng et al. (2005)
Eomylus borealis			■					Chow and Qi (1978), Dashzeveg and Russell (1988), Meng et al. (1998)
Eomylus zhigdenensis			■	■				Dashzeveg and Russell (1988), Dashzeveg et al. (1998)
Eurymylus laticeps			■					Matthew and Granger (1925), Sych (1971), Dashzeveg and Russell (1988)
Hanomys malcolmi			■					Huang et al. (2004)
Heomys orientalis	■	■						Li (1977), Wang et al. (2016)
Khaychina elongata			■					Dashzeveg and Russell (1988)
Mimotona lii	■							Li (1977), Dashzeveg and Russell (1988)
Mimotona robusta		■						Li (1977)
Mimotona wana	■	■						Li (1977), Li and Ting (1985, 1993)
Mina hui	■	■						Li et al. (2016)
Palaeomylus lii			■					Meng et al. (2005)
Sinomylus zhaii			■					McKenna and Meng (2001)
Taizimylus tongi			■					Mao et al. (2017)
'Rodentiaformes'								
Alagomys russelli			■					Dawson and Beard (1996)
Alagomys inopinatus				■				Dashzeveg, (1990)
Alagomys oriensis				■				Tong and Dawson (1995), Li (2016)
Neimengomys qii			■					Meng et al. (2007)
Tribosphenomys minutus			■					Meng et al. (1994, 1998, 2007), Meng and Wyss (2001)
Tribosphenomys secundus			■					Lopatin and Averianov (2004a), Meng et al. (2007)
Tribosphenomys tertius			■					Lopatin and Averianov (2004b)
'Ischyromyidae'								
Paramys adamus			■					Dawson and Beard (1996)
Paramys taurus			■					Rose (1981), Ivy (1990), Anderson (2008)
Franimys amherstensis			■					Rose (1981), Ivy (1990), Anderson (2008)
Acritoparamys atavus			■					Rose (1981), Ivy (1990), Anderson (2008)
Ischyromys douglassi					■			Rankin et al. (2020), Emry (1981), Black (1968), Wood (1974), Heaton (1996)
Ischyromys blacki					■	■		Wood (1974), Heaton (1996)
Ischyromys veterior						■		Wood (1976), Wood (1980), Heaton (1996)
Ischyromys typus							■	Heaton (1993)

Table 1: Basal Glires, 'Rodentiaformes' and members of 'Ischryomyidae' mentioned in the text. Fossil ages and justification for ages are also provided. This table is adapted from Fostowicz-Frelik (2020: their Table 5.1) and includes additional species and information for *Alagomys,* as well as the addition of *Ischyromys*.

et al., 2007), and *Alagomys* from North America and Inner Mongolia (Dawson and Beard 1966; Dashzeveg, 1990). The paraphyletic group 'Ischyromyidae' is much more speciose and contains many different genera: *Acritoparamys*, *Franimys*, *Ischyromys* and *Paramys*, to name a few (Korth, 1994). Ischyromyidae are distinguished from Rodentiaformes by their elongated bodies, larger size, and were in general quite a bit sturdier than their gracile relatives. The geographic distribution of Ischyromyidae is another dissimilarity between the two groups. 'Ischyromyid' fossils are exclusively found in North America, despite living contemporaneously with the Asian Rodentiaformes and some later eurymylids. The disjunct distribution between North American Ischyromyidae and their presumed Glires ancestors in Asia has been termed the 'Paleocene Paradox' (Dawson, 2015; Fostowicz-Frelik, 2020). Unfortunately, the phylogenetic relationships of Ischyromyidae to other rodents, and between ischyromyids themselves, are not very clear. Ischyromyids have been interpreted as many things, including primitive Rodentia (Flynn, 2008), a sister group to Rodentia (Dawson and Beard, 1996), paraphyletic with species on the stem of Rodentia and others in 'true' Rodentia (Meng and Wyss, 2001), and finally as a paraphyletic group with some species within crown Sciuromorpha and others on the rodent stem (Asher et al., 2019).

The sheer quantity of Glires fossil evidence from Paleocene Asia points towards an Asian origin for Glires, and within it, Lagomorpha. The origins of Rodentia are also likely to lie within Asia. However, Glires known from localities on both sides of the Pacific, such as the rodentiaform *Alagomys*, obscure matters. If an Asian origin of rodents is likely, then we would expect that Asian specimens of *Alagomys* would pre-date North American specimens. However, the American Clarkforkian (Late Paleocene) fossils of *Alagomys* pre-date the Asian Bumbanian (Late Eocene) *Alagomys* specimens by millions of years (Dashzeveg, 1990; Tong and Dawson, 1995; Bowen et al., 2012). The boundary between the Gashatan age and Bumbanian age is often debated, although they are usually ascribed to the late Paleocene (Clarkforkian NALMA) and early Eocene (Wasatchian NALMA) respectively (Dashzeveg, 1988; Beard and Dawson, 1999). A study by Bowen et al. (2012) which combined isotope stratigraphy and magnetostratigraphy, alongside biochronology, suggested a Gashatan-Bumbanian boundary between 55.7–54.97 Ma, which is remarkably close to estimates for the Paleocene-Eocence boundary (55.72–55.96 Ma; Charles et al., 2011) and would suggest that Clarkforkian *Alagomys* do indeed predate their Asian Bumbanian neighbours. Once again, the Paleocene Paradox muddies the waters of biogeographic history. In addition to this, the uncertain relationships of North American Ischyromyidae, some of which are of comparable age to many ancient Asian taxa, leave open the possibility that North America could play a nuanced role. This is especially relevant for Sciuromorpha, which may be closely related to ischyromyids or hold some of them within its crown. Both Sloan (1969) and Wood (1977) proposed a North American origin for Rodentia based on the presence of Paleocene

and Eocene members of Ischyromyidae. An Asian origin of Glires seems likely. For at least Sciuromorpha, the North American continent merits consideration and will be investigated as part of my study.

Myomorpha, also referred to as the mouse-related-clade (Fabre et al., 2015) or Supramyomorpha (Elia et al., 2019), may also have a strong case for North American roots. Numerous Eocene fossils that possess ancestral myomorph characters (loss of premolars and enlarged infraorbital foramen), can be found in both Asia and North America. On the Asian side there are species such as *Aksyiromys*, *Blentosomys* and *Ulkenulastomys* from eastern Kazakhstan (Shevyreva, 1984) and *Palasiomys*, *Raricricetodon* and *Pappiocricetodon* from China (Tong, 1997). Eocene myomorphs from North America include *Simimys* (Wilson, 1935), *Nanomys* (Emry and Dawson, 1972), *Armintomys* (Dawson et al., 1990), *Heliscomys* (Cope, 1873) and *Elymys* (Emry and Korth, 1989). The fossil *Elymys* is of interest as it is sometimes argued to be the earliest myodont (Muroidea + Dipodoidea) or at least the earliest dipodoid fossil (Rodriguez et al, 2009). With this in mind, Emry (2007) suggested that Myodonta originated in North America, after which *Elymys*-like taxa could have made the journey to Asia in the middle Eocene. These taxa then presumably diversified in Asia before cricetid and dipodid rodents travelled back to North America in the late Eocene and throughout the Miocene (Emry, 2007). Apart from Rodriguez et al's (2009) study on dental characters, there have been no other studies to include *Elymys* in phylogenetic analysis. However, the placement of this fossil is likely to be key in determining whether or not Myomorpha has origins in North America or Asia, something my study aims to address.

The arguments against a North American origin for any rodent suborder were strongly put forward by Beard (1998). His 'East of Eden' model took into account the prevalence of Glires fossil taxa in Asia as well as the idea that immigration from a larger landmass (Asia) to a smaller landmass (North America) would be more likely than the reverse. From Beard's point of view, North America was a "cul de sac" (1998, p. 1) of Asia at the time, with most traffic across the Bering Land Bridge coming from the East. Yet for alagomyid rodents, dispersal appears to be *to* Asia rather than from it. The Bering Land Bridge had been a potential route of dispersal between the two continents until the Pliocene and very little land connection (if any) has joined the continents across the north Atlantic since the earliest Eocene (Jiang et al., 2019). In theory, there has been ample opportunity for travel in both directions for an extended period of time. Jiang et al. (2019) examined rates of biotic interchange across the land bridge for animals, plants and fungi, and found that this opportunity was not fully taken advantage of. With all biota taken into account, the number of dispersal events from Asia to North America outweighed dispersal from the other direction throughout the Cenozoic. While the rates of dispersal from Asia to North America have always remained higher, there was a peak in movement from North America to

Asia at the end of the Eocene (40–34 Ma), which coincided with global cooling (Jiang et al., 2019). However, if we examine Jiang et al's (2019) data for mammals specifically, there is no peak in mammal dispersal at this time. In fact, the number of mammal dispersal events sharply increase later on during the Oligocene Cooling Event (30 Ma), and movement of taxa from North America suddenly decreases at this time. For mammals as a whole, immigration seems to have been predominantly from Asia to North America and partly driven by climatic factors.

The effects of Asian climate and uplift of the Tibetan Plateau

The role of climate in mammal dispersal to and from Asia is clearly of importance and could explain the patterns we see in the rodent fossil record. It would be impossible to consider Asian climate or immigration at the time of Glires diversification without also discussing the impact of one important tectonic event: the collision of India and the resulting uplift of the Tibetan Plateau. The Tibetan Plateau (also known in China as the Qinghai-Tibet Plateau or Qing-Zhang Plateau) is appropriately referred to as 'the roof of the world'. It is the highest and largest geological monument of its kind, spanning 2.5 million km^2 and reaching 4,500m above sea level (Wang et al., 2015). Argand (1922) was the first to suggest that the plateau arose from the collision between an Indian plate with Eurasia. This theory is widely accepted today, but the exact timing and mechanisms are still being investigated. It is thought that the Indian landmass started to peel away from Madagascar around 140 Ma, before breaking off around ca. 83 Ma and then accelerating towards Asia (Torsvik and Cocks, 2016). The first sign of elevation in Central Asia occurred during the Cretaceous before the collision took place (Figure 18). As the Indian Plate approached and the Tethys ocean floor began to subduct, the crust in what is now central Tibet started to thicken, as well as fold and fault, resulting in some areas rising above sea level by the late Cretaceous (Royden et al., 2008). The Indian landmass then made contact with the Asian continent during the early Eocene, ca. 50 Ma (Torsvik and Cocks, 2016). It was the subduction of the Indian continental crust and mantle lithosphere under the Asian continent that led to the Himalayan orogeny. The impressively tall Tien Shan, Karakorum, Himalaya and Altai mountain ranges that we see today were caused by a second peak of uplift in the Late Paleogene-Early Neogene, and they are still growing today (Torsvik and Cocks, 2016). The Tibetan Plateau is a very large area and it was pointed out by Wang et al. (2011) that the various regions of the Tibetan Plateau experienced uplift during different time periods. This complicates our understanding of the timing of these events, but to overcome this we can identify overall trends in uplift. Wang et al. (2011) were able to identify the overlapping periods of uplift and extracted the contemporaneous episodes: 60–35 Ma (Middle Paleocene to Late Eocene), 25–17 Ma (Late Oligocene to Early Miocene), 12–8 Ma (Middle to latest Miocene) and since 5 Ma (Pliocene onwards).

Given its tectonic activity, size and elevation, the Tibetan Plateau is thought to play a significant role in weather patterns across Asia. The vast surface area and elevation of the plateau causes it to absorb a far greater quantity of solar radiation than land at sea level (Lin et al., 2018). The air above the plateau is therefore at a higher temperature than the surrounding terrain and ocean. Like in all coastal systems, the difference in temperature between land and sea creates areas of low pressure inland. The warm air over land rises and creates pockets of low pressure underneath it. These pockets of low pressure draw in moist ocean air. Any coastal air that is blown towards a mountain is forced upwards, but cooler temperatures at altitude cause a decrease in air volume and so the moisture precipitates as rain. The elevation of the plateau exacerbates this effect and was hypothesised by Flohn (1957) to cause summer deluges (monsoons) in India. However, further studies have identified an alternative role of the plateau in driving monsoons. Boos and Kuang (2013) argue that the Himalayas are the main heat source for monsoons, not the plateau itself, with the latter forming a barrier to colder northern winds. Either way, it is generally accepted that South Asian monsoons as we know them today began during the late Oligocene, with monsoon-like weather appearing as early as the late Eocene (Sun and Wang, 2005). Climate models have indicated that the uplift of the Tibetan Plateau and the reduction of the Paratethys Sea, which are both linked to the Indo-Asian collision, resulted in monsoon intensification as well as increased levels of erosion (Dupont-Nivet et al., 2007). The inevitable aridification that followed this is linked to the onset of the Eocene-Oligocene transition, a rapid cooling of the planet 34 million years ago (Dupont-Nivet et al., 2007). It is debated if this activity was the major driver behind the Eocene-Oligocene climate transition, but it seems likely that the Tibetan Plateau played a role. Weather patterns across Eurasia today are highly dependent on Tibetan Plateau processes. The amount of snow that falls in Eurasia is correlated with the Indian monsoon, and snow cover on the plateau itself is known to lengthen or cut short the duration of summer in monsoon affected areas (Hahn and Manabe, 1975; Liu et al., 2004).

There is good reason to believe that uplift of the Tibetan Plateau could have had the potential to affect the diversification of animal groups. It has remained a popular idea that increased diversity and speciation occurs primarily under tough environmental conditions, as opposed to in tropical refugia where physical/environmental forcing is not as important a driver of evolution (Fortelius et al., 2014). More specifically, Badgley (2010) hypothesised that severe tectonic activity would coincide with peaks of species diversification, as increased speciation can result from topographic complexity. We might therefore expect Glires on the Asian continent to show peaks of diversity that coincide with periods of major uplift (such as those described by Wang et al., 2011) in the Tibet area. Wang et al. (2015) present several cases where the combined forces of climate and uplift of the Tibetan Plateau are hypothesised to have influenced vertebrate evolution during the Cenozoic. They set out a case for the

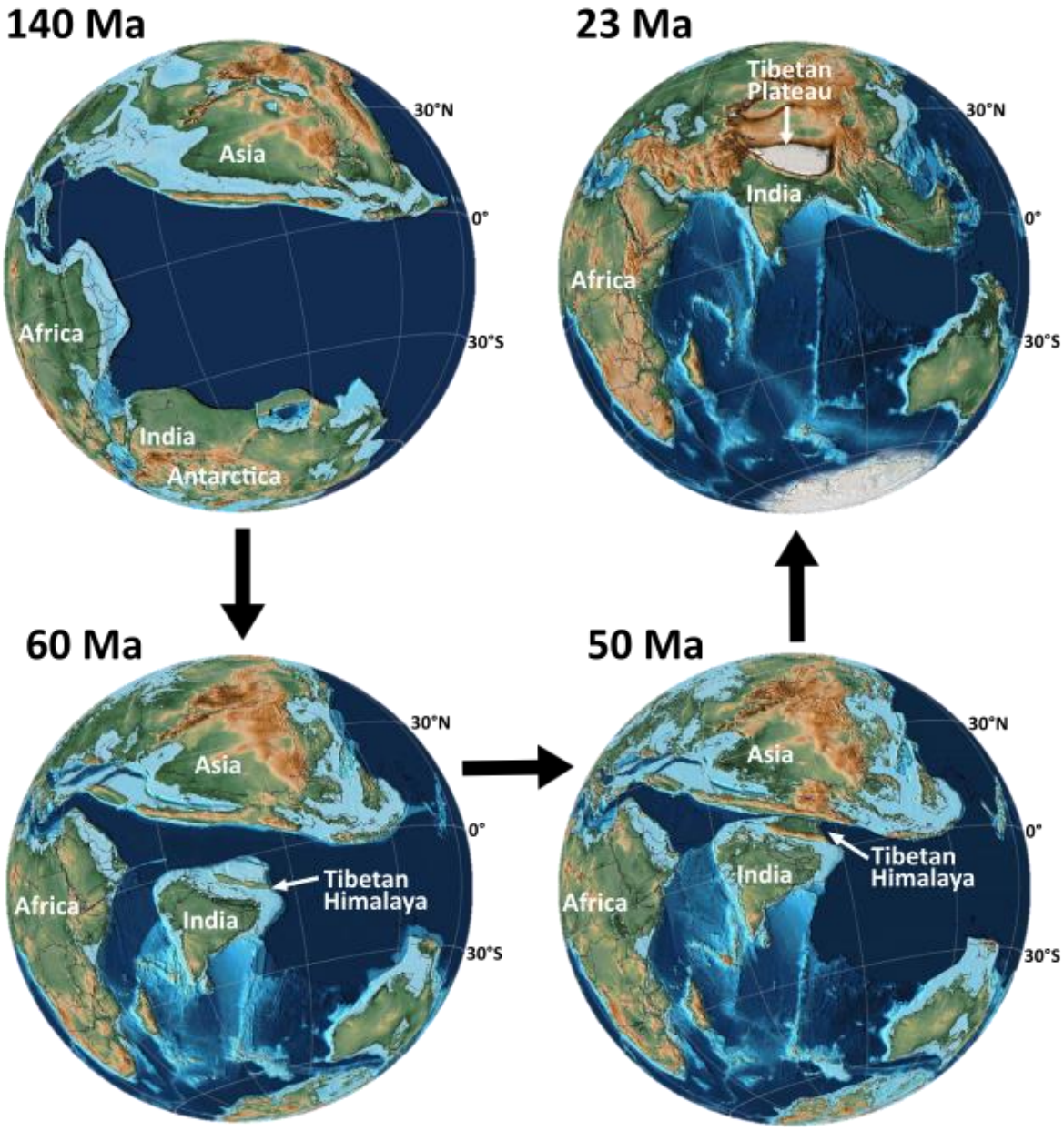


Figure 18: Movement of the Indian plate northwards and collision with Southeast Asia. Palaeomaps are sourced from Scotese (2016).

plateau acting as a refugia for cold-adapted species (including the Tibetan woolly rhinoceros, snow leopard and artic fox) during the various climatic cycles of the Ice Age, as well as populating the broader continent with lineages that were pre-adapted to cold conditions before the onset of the Ice Age. A compelling case can be made for Ice Age refugia and it would be interesting to consider the possibility of Asian refugia for Glires across the whole of the Cenozoic. If such refugia were to exist, we might expect dips in Glires diversity to be followed by clades arising in the Tibetan Plateau, Himalayas or surrounding areas. The complexity and uncertainty regarding the exact timing of events in the Tibetan Plateau, due to its vast size and the geological record, could make it hard to draw solid conclusions about the relationship between uplift intensity and animal diversity. However, the contemporaneous periods of uplift calculated by Wang et al. (2011) are good indicators of overall trends in uplift across the entire region and are our best approximation of uplift timetable. That being said, it would be sensible to treat any relationship found between overall diversity and uplift trends tentatively until a more resolved record can be obtained.

Ctenohystrica and dispersal westwards

While Sciuromorpha and Myomorpha have complicated histories in Asia, there is one rodent group that has largely abandoned the continent. Living members of Ctenohystrica are found in Africa and South America, with two notable exceptions: the Laoatian rock rat (*Laonastes aenigmamus*) from Southeast Asia and eight species from the family Hystricidae ('old world hystricognaths' or 'phiomorphs'). The most diverse group within Ctenohystrica is the monophyletic clade Hystricognathi. These animals possess 'anatomical hystricognathy' whereby the angular process lies laterally to the incisor alveolus (Tullberg, 1899). It is now widely accepted that Hystricognathi share a common ancestor with Ctenodactylomorphi (Elia et al, 2019), a group which consists of gundis in Africa (Ctenodactylidae) and the Laoatian rock rat (Diatomyidae) in Asia. 'Ctenodactyloidea', the fossil taxa considered more similar to Ctenodactylomorphi than Hystricognathi (Marivaux and Boivin, 2019), were restricted to Asia for the entirety of the Palaeogene. It would therefore seem likely that Hystricognathi originated in Asia as well. However, the oldest anatomically hystricognathous rodents appear in the Late Middle Eocene of South America (stem caviomorphs, e.g., *Canaanimys* and *Cachiyacuy*) and Africa (stem 'phiomorph' *Protophiomys*) simultaneously (Antoine et al, 2012; Marivaux et al., 2014; Boivin, 2017; Boivin et al., 2019). Anatomical hystricognaths are not found in Asia until a few million years later in the Late Eocene, and so the question is, did Hystricognathi originate in South America, Africa or Asia?

Phylogenies based on dental data that were retrieved by Marivaux and Boivin (2019) indicate multiple dispersal events of hystricognathous rodents from Asia to Africa and possibly back again, as well as implying possible Hystricognathi ghost lineages extending into the Middlle Eocene for Asian and

African taxa. In regards to the colonisation of South America, molecular clock analyses by Poux et al. (2006) suggest that caviomorphs split from old world hystricognaths 45.4 to 43.7 Ma, possibly a few million years before they immigrated to South America between 45.4–36.7 Ma. A more recent study by Antoine et al. (2012) narrows this dispersal window down to 42–43 Ma. Just one million years later ca. 41-42 Ma, the global temperature peaked during the Middle Eocene Climate Optimum, suggesting that rodents travelled to South America when environments were warm and wet (Zachos et al., 2008; Antoine et al., 2019). As discussed by Marivaux and Boivin (2019), the origins and biogeographic history of Ctenohystrica likely depend on the discovery of new material, interpretation of stem-groups and their phylogenetic affinities. The latter will be investigated in my study.

A total-evidence approach

Since our interpretation of Glires evolutionary history relies upon the interpretation of fossil data, it is important that we find the best way to incorporate this information with the molecular record. Fossils can be placed *post hoc* on a phylogeny based on molecular data, or fossils can be analysed with a backbone constraint based on said molecular data. The total-evidence approach considers several types of data simultaneously, with some level of data interaction between the partitions allowed. Considering many sources of data in tandem has the benefits of taking all phylogenetic signal into account when building a tree. Signal that is common to each partition, but weak when considered alone, can be amplified by including data from lots of different sources (Gatesy et al., 1999; Gatesy and Arctander, 2000; Gatesy and Baker, 2005). This 'hidden signal' can only be revealed by using such methods. The phylogenetic positions of fossils are known to be influenced by the addition of molecular data, and morphological data from fossils can also change the affinities of living taxa (Asher et al., 2005; Wiens et al., 2010; Thompson et al., 2012). In its 'tip-dating' form, fossil taxa in a total-evidence context help define the age of a clade during the tree building process itself (Ronquist et al., 2012). Combining molecular and fossil data therefore has benefits over analysing these data partitions separately.

A recent development in the study of Glires has been the publication of ultraconserved elements (UCEs). These are short pieces of non-coding DNA that are shared by genetically distant taxa and can be captured in their thousands using target enrichment processes (Bejerano et al., 2004). UCEs are highly conserved and so are thought to be less prone to saturation than some other types of molecular data (Faircloth et al., 2012). This in turn could make UCEs ideal for helping to resolve relationships through deep-time. In the case of Glires, UCEs suggest different phylogenetic affinities at the base of Rodentia than other types of data. Most molecular studies support a Myomorpha + Ctenohystrica sister clade to Sciuromorpha (Churakov et al., 2010; Fabre et al., 2012; Asher et al., 2019). However, phylogenies based on UCEs propose a Ctenohystrica + Sciuromorpha sister clade to Myomorpha

(Esselstyn et al., 2017; Swanson et al., 2019). Knowing the basal relationships of Glires is vital to understanding their diversification and historical biogeography. Yet so far, no analyses of Glires have combined fossil and UCE data under a total-evidence framework. Swanson et al. (2019) gathered the largest dataset of UCEs for Glires species, but grafted their fossil taxa onto the topology post-phylogenetic analysis. Only two studies have used total-evidence approaches to analyse fossil and UCE data simultaneously, in birds (Chen et al., 2019) and iguanas (Scarpetta, 2020), and so part of my study aims to address this gap in the literature for Glires.

Aims and hypotheses

The aims of this chapter are to use total-evidence analyses of fossil and molecular data to investigate the timing and origins of Glires diversity. By including key fossils from across the world, I hope to determine the ancestral ranges of each Glires group. In particular, the inclusion of older North American and Asian fossil taxa should help to trace the origins of Myomorpha and Sciuromorpha. I posit that 'ischyromyids' and Sciuromorpha will share a close history, and so a North American origin for this clade is likely. I also expect that if the fossil *Elymys* is found to be a relatively basal member of Myomorpha, then this clade would also have origins in North America, challenging Beard's idea that the continent acted only as a cul de sac of Asia. My study also aims to trace the roots of Ctenohystrica. Despite having very little representation nowadays in Asia, there are many fossil 'ctenodactyloid' taxa from the continent. The series of dispersal events within Ctenohystrica is not clear, but I predict that at least one dispersal event from Asia to Africa took place. My predictions on biogeography could. of course, be heavily influenced by the inclusion or exclusion of UCEs in my analysis, as I anticipate this to change the relationships of both living and extinct forms of Glires. Finally, my study also aims to investigate if the uplift of the Tibetan Plateau played a role in observed patterns of Glires diversification. If the Tibetan Plateau acted as a refuge or training ground for cold-adapted Glires species, then we might expect to see peaks in Asian Glires diversity coinciding with cooler global temperatures. The tectonic upheaval of the Tibetan Plateau and its supposed influence on climate may also result in peaks of diversification during major periods of uplift.

Methods

Morphology plus protein coding DNA matrix assembly

In order to perform the phylogenetic analysis, I assembled a matrix of morphological and molecular characters. I utilised the data matrix from Asher et al. (2019), which comprises of 14 protein coding genes: eight nuclear (ATP7A, A2AB, IRBP, RAG1, VWF, CNR1, BRCA1 and GHR) and six mitochondrial

(12S, 16S, CYTB, CO1, CO2 and CO3), totalling 15,407 nucleotides. I then added one additional genus to the matrix, *Cricetulus,* the sequences for which were sourced on GenBank. Details of the sequences sourced for each genus in Asher et al. (2019) and *Cricetulus* can be found in Appendix 1b. I aligned the *Cricetulus* sequences to Asher at al's (2019) matrix using an opalescent algorithm within Mesquite version 3.6 (Wheeler and Kececioglu, 2007, Maddinson and Maddinson, 2019) and then made small manual adjustments by eye. I partitioned each locus by codon position before using PartitionFinder 2 (Lanfear, 2017) to find the best-fitting partition schemes and models of evolution, based on the Bayesian Information Criterion. The models used for each partition can be found in Appendix 3a.

I included the morphology data of Asher et al. (2019), which consists of 219 hard tissue characters for 60 extant taxa (52 Glires) and 42 fossils (27 Glires and stem Glires). The morphological characters in Asher et al. (2019) are based on those used in Meng et al. (2003), Asher et al. (2005) and Wible (2007). I coded the characters of an additional 21 fossil Glires taxa: *Alagomys, Birbalomys, Branisamys, Dawsonolagus, Elymys, Eocardia, Eogliravus, Eutypomys, Exmus, Franimys, Gaudeamus, Hulgana, Knightomys, Paradelomys, Prolapsus, Protechimys, Sallamys, Sayimys, Tamquammys, Tarnomys* and *Tsinglomys.* These were chosen for their completeness, biogeographical relevance, and to maximise taxonomic coverage. I obtained microCT scans of *Cricetulus*, *Franimys, Ischyromys, Paleolagus* and *Tarnomys,* and also examined the microCT scans of *Gomphos*, *Ischryomys* and *Paradelomys* from Rankin et al. (2020). I prepared virtual reconstructions of the microCT scans using the open source software Drishti v.2.6.4 (Limaye, 2012). Other taxa were examined through a combination of observing museum specimens and images found within the literature. A list of the specimens I examined, any changes to the character coding of Asher et al. (2019), and literature consulted, can be found in Appendix 1a. A matrix of my coded morphological characters can be found in Appendix 1c.

Finally, I concatenated my morphology and protein coding loci partitions to make one supermatrix, which can be found in Appendix 3a. Hereafter, I refer to this alignment as my 'morph+coding' dataset.

Ultraconserved element (UCE) matrix assembly

I sourced UCE data from three published datasets: Swanson et al. (2019), Esselstyn et al. (2017) and Faircloth et al. (2012). Thirty-two sequences of concatenated UCE loci matched the taxa present within my morph+coding dataset: 25 genera from Swanson et al. (2019), 6 from Esselstyn et al. (2017) and 1 from Faircloth et al. (2012). Each source of UCEs used the same 5,060 UCE probe set of bait sequences to harvest UCEs from mammal genomes. However, because each paper varies in the downstream processing of the captured UCEs, not every UCE harvested was necessarily included within their respective final alignments. Furthermore, the order in which homologous UCEs appeared within the concatenated sequences differs between the three alignments. Given that the majority of rodent

sequences were provided by Swanson et al. (2019), I decided to extract individual UCEs from the other two datasets and re-align them to the concatenated sequences of Swanson et al. (2019). First of all, I performed a local alignment using the blastn algorithm (Johnson et al., 2008), to query the relevant sequences from Esselstyn et al. (2017) and Faircloth et al. (2012) against the Swanson et al. (2019) alignment. I then scrutinised the list of hits by eye and removed any spurious sequences using the Python script 'faSomeRecords.py' (Kent, 2014). Once extracted, I then rearranged the UCEs to reflect the order of appearance in which they occur within the Swanson et al. (2019) alignment using SAMtools (Li et al., 2009). I took the newly rearranged UCES from Esselstyn et al. (2017) and Faircloth et al. (2012), concatenated them and then aligned them to the Swanson et al. (2019) alignment using Mafft 3.709 (Katoh et al., 2002). As the amount of data to align was very large, I accessed Mafft through the Cambridge Service for Data Driven Discovery's High Performance Computing cluster (CSD3 HPC). The following non-default options were used: 'add' allows new sequences to be aligned to an existing alignment, 'reorder' allows sequences to be ordered according to similarity, 'memsave' allows a linear-space DP algorithm to be used (less memory is used at the expense of speed) and 'thread' -1, which automatically chooses the number of threads to run based on the number of computer cores available. I viewed the alignment of UCEs and made small manual adjustments using Mesquite version 3.6 (Maddinson and Maddinson, 2019). The final alignment length of the UCEs amounted to 914,885 nucleotides from 2,213 UCE loci. I used the blastn algorithm once more to query the UCEs against my morph+coding matrix and to make sure that there was no overlap between the UCEs and my protein coding genes. Finally, I concatenated the UCEs with my morph+coding matrix to make a larger dataset, hereafter known as my 'UCE dataset', which can be found in Appendix 3b. I took the decision not to run any analyses to find partition schemes and substitution models for the UCEs for two reasons. Firstly, given the length of the alignment and number of taxa, this would have been a very computationally intensive task (Esselstyn et al., 2017). Secondly, studies that include UCE data have found that single partition datasets recover the same topology as multiple partition models and may reach convergence more easily (McCormack, 2011; Tarver, 2016; Sato et al., 2019).

Bayesian tip-dating analyses

For the phylogenetic reconstruction based on my morph+coding dataset, I adopted a Bayesian tip-dating approach, which estimates phylogeny and divergence times concurrently (Ronquist et al., 2012). This method has the advantage of incorporating fossil data during the analysis itself whilst taking uncertainty into account under a Bayesian probabilistic approach. I used MrBayes 3.2.7 (Ronquist and Huelsenbeck, 2003; Altekar et al., 2004) for this purpose and ran the parallel version through the CSD3 HPC. I implemented the partitioning scheme and models of evolution selected by PartitionFinder 2 for the alignment of protein coding molecular data. For the morphological partition,

I opted to use a Markov *k*-states model (Mkv; Lewis, 2001), which uses gamma-shaped rate variation. There are arguments for and against the use of the Mkv model over more traditional parsimony approaches. The 'no-common-mechanism' of parsimony means that an independent parameter is assigned to each branch length for each character and so is computationally less parsimonious than the Mkv common-mechanism model. Some authors argue that the common-mechanism implied by the Mkv model fits molecular data quite well, but is not realistic for morphological data as it evolves in a different way (Goloboff et al., 2019). However, other studies have found that the Mkv model produces more accurate, albeit less precise, trees than both equal and implied-weights parsimony (O'Reilly et al., 2016; Puttick et al., 2018). In this instance, I chose accuracy over precision on account of my downstream analyses. Importantly, I configured the rate prior to allow rates to vary between my morphological partition and each of my molecular partitions. For the molecular clock, I chose an independent gamma rate relaxed-clock model to account for evolutionary rate variation across branches over time. For each fossil taxon in my matrix, I specified a uniform prior corresponding to the minimum and maximum ages for that taxon based on records held within the Paleobiology Database (PBDB). I gave extant taxa a fixed age of zero. In order to enable these calibrations, I implemented the fossilized birth-death sampling process. I treated fossil taxa as tips with no descendants. All commands given to MrBayes can be found within Appendix 3a. I ran two independent runs of four Markov Chain Monte Carlo (MCMC) chains (three cool and one heated) for 120,000,000 generations and sampled the chains every 1,000 generations. To test for convergence, I examined the standard deviation of mean split of frequencies and the potential scale reduction factor. I also examined the effective sample sizes using TRACER v.1.7.1 (Rambaut et al., 2018). After examining the MCMC output, I discarded the first 20% of trees as 'burn-in' and then summarised the remaining trees to create an 'allcompat' consensus tree, where all compatible groups are added to the 50% majority rule consensus.

Maximum Likelihood analyses

Given the large size of the UCE dataset and computational time constraints, I did not include this data in my Bayesian tip-dated analyses. Initial attempts at Bayesian tip-dating analyses for the UCE dataset estimated runtimes that were not feasible within the scope of this PhD, a common problem for UCE and other large phylogenomic datasets in general. Where Bayesian inference has been used to estimate phylogeny with UCEs, compromises are typically made by reducing the number of taxa or loci (Blaimer et al., 2018; Borowiec, 2019). Instead of dataset reduction, most other studies adopt a Maximum Likelihood approach when using UCEs (Zhang et al., 2019). Maximum Likelihood has the advantage of being less computationally intensive and so is much faster than Bayesian inference when datasets are very large. The frequentist approach to probability however, does not incorporate priors

of any kind and so usually, I would prefer the method which takes all of the information we have into account. However in this case, reducing the size of my dataset would effectively be ignoring information in much the same way, but with little benefit. For my own phylogenetic analysis of the UCE dataset, I decided to use the program RAxML (Stamatakis, 2014), which contains many options for parallel process optimization. I ran the RAxML-PTHREADS-AVX2 binary, accessed via raxmlGUI 1.5 beta (Silvestro, 2012), across four CPUs on a local computer. Currently, there is no way to 'checkpoint' the maximum likelihood search process, and so the CSD3 HPC, although perhaps faster, could not be used due to run time limits. I implemented the same partitions schemes suggested by ParitionFinder 2 as used above. As RAxML can only implement one model of rate heterogeneity across partitions, I chose the GTR+G+I model as it is the one suggested for most of my protein coding DNA partitions by PartitionFinder 2. However, I was able to configure the run to calculate individual per-partition branch lengths. All of the commands I gave to RAxML can be found in the RAxML block at the end my UCE alignment in Appendix 3b. I ran a Maximum Likelihood search with 100 rapid bootstraps.

In order to convert the best tree produced by RAxML into a dated ultrametric phylogeny, I carried out node dating using a penalized likelihood approach via TreePL (Smith and O'Meara, 2012). I used 13 fossil calibrations, including nine used by Swanson et al. (2019) for their rodent UCE dataset, and a further four calibrations from Benton et al. (2015). The calibrations describe the soft maximum dates for a fossil, extending to the start of the geological epoch prior to the fossil's first occurrence as well as hard minimum dates. The *Mus-Rattus* divergence time however, is taken from Kimura et al. (2015). My calibrations and parameters given to TreePL can be found in Appendix 3c.

Diversity and environmental measures

In order to comment on the changes in Glires diversity over time, I calculated values of taxonomic richness. Firstly, I downloaded all occurrence records of fossil Glires from the PBDB (paleobiodb.org) and then cleaned the data. The raw counts of genera cannot be used as an indicator of true diversity as it is widely recognised that the fossil record is subject to many different types of bias (Raup, 1972). Generally, studies compensate for incompleteness in the fossil record by standardizing their data and calculating estimates of diversity. Quorum subsampling (SQS; Alroy, 2017) is the most common method and uses generic rarefaction to provide relative estimates of diversity. This approach works well for fossil data, but newer methods are now capable of providing accurate and absolute measures of richness. Alroy's (2018) 'extrapolated squares' is one such method and has been shown to handle datasets skewed towards rare species (a common occurrence in palaeontology) more effectively than SQS. Absolute measures of diversity are more intuitive for my purposes and so I applied the squares method to the Glires raw data. To calculate extrapolated squares, I followed an R script shared by Bethany Allen (University of Leeds), which contains Alroy's (2018) squares formula, which can be

found in Appendix 3d. My cleaned occurrence data and extrapolated squares estimates of diversity can be found in Appendix 3e.

In addition to estimates of diversity, I examined two potential factors that could provide insight into the environmental pressures imposed upon Glires during their evolution: global temperature proxies and uplift of the Tibetan Plateau. Firstly, I gathered data on levels of deep-sea oxygen ($\delta^{18}O$) isotopes over time. The proportion of ^{18}O to ^{16}O present in the carbonate component of foraminifera can be used to estimate temperature, as decreases in heavy oxygen are associated with increasing heat (Veizer, 1999). I combined $\delta^{18}O$ data (Appendix 3f) from Veizer (1999) and Zachos et al. (2001) and applied a five-point rolling mean to make the trends over time easily visible.

The second factor I examined alongside diversity is uplift of the Tibetan Plateau. The elevation of the Tibetan Plateau over time is not as easily measured as $\delta^{18}O$, and currently we do not have a continuous measure of the plateau's height through deep-time. In order to examine Glires diversity alongside changes to the Tibetan Plateau, I incorporated Wang et al's (2011) episodes of major tectonic uplift. Tectonic uplift is measured by structural deformation, thermo-chronology and sedimentary deposit data, although these measures have indicated that movement was not contemporaneous across the different regions of the Tibetan Plateau (Wang et al., 2011). However, according to Wang et al. (2011), the main intervals of overlap occur between 60–35 Ma, 25–17 Ma, 12–8 Ma and from 5 Ma. In order to test whether or not diversity levels are significantly correlated with Tibetan Plateau activity or temperature estimates, I carried out Spearman's correlation tests.

Biogeographical analyses

Finally, I investigated the biogeography of Glires through time by reconstructing ancestral distributions on phylogenies based on my morph+coding and UCE datasets. Like general ancestral state reconstruction, there are options for both probabilistic methods, such as likelihood and Bayesian, as well as non-probabilistic approaches such as parsimony. I decided to test probabilistic models using the popular R scripts that form the package BioGeoBEARS (Matzke, 2013), which I accessed through RASP 4.2 (Yu et al., 2015). The R scripts require each taxon on the phylogeny to be allocated states relating to the distribution of that taxon. There are many ways to define geographical areas for biogeographic analyses, but I chose to split the map into six simple areas: Australasia, Africa, Asia, Europe, North America and South America. Given the timescale of Glires evolution, it is hard to choose meaningful geographic areas which remain relevant over time as landmasses shift. I coded the geographical states of living taxa based on their current distributions, but did not include areas where a taxon is considered a recent introduction. For the fossil taxa, I coded their distributions based on the information provided with specimens and in PBDB occurrence records. The models used by

BioGeoBEARS require that trees are completely bifurcating and so I made small adjustments to my topology based on the morph+coding dataset. I resolved taxa according to the literature, whilst remaining as close as possible to the original topology provided by MrBayes. I resolved the position of my *Eogliravus* + *Knightomys* clade to the stem of Sciuromorpha, based on Marivaux et al. (2016) and Storch and Seiffert (2007), who both consider *Eogliravus* to be a primitive glirid (Gliridae). I resolved my *Sciuravus* + *Birbalomys* clade to the stem of crown Rodentia based on the position of *Sciuravus* in Asher et al. (2019). I then tested six models of ancestral distribution on my resolved morph+coding matrix topology and my UCE matrix topology. Based on the AICc scores, the best fitting model of ancestral area reconstruction for the morph+coding matrix topology was DIVALIKE+J (Dispersal–Vicariance Analysis; Ronquist, 1997). The DIVALIKE model is a Maximum Likelihood interpretation of Ronquist's (1977) Dispersal-Vicariance-Analysis. The J parameter describes cladogenetic dispersal, whereby speciation can occur in a different geographic area from the locations of the ancestor. The best fitting model for the UCE dataset tree was BAYAREALIKE+J. The BAYAREALIKE model is a Maximum Likelihood interpretation of Landis et al's (2013) Bayesian inference of historical biogeography for discrete areas analysis.

Results

Total-evidence phylogenies and topological differences

Figure 19 shows the tip-dated phylogeny based on my morph+coding dataset. This tree suggests that Lagomorpha and Rodentia split 71 Ma in the late Upper Cretaceous. The eurymylid taxa form a group on the stem of Rodentia and the mimotonids are clustered around the stem of Lagomorpha. Eurymylidae is not a monophyletic group however, as *Sinomylus* is positioned on the base of the Lagomorpha stem. The alagomyid rodentiaformes (*Alagomys* and *Tribosphenomys*) do form a monophyletic group, on the rodent stem. The morph+coding-based topology also supports the monophyly of three rodent crown groups (Ctenohystrica, Myomorpha and Sciuromorpha), with Sciuromorpha comprising the root. The tree implies that Sciuromorpha split from Ctenohystrica-Myomorpha ca. 65 Ma during the earliest Paleocene and then three million years later, Ctenohystrica and Myomorpha split from each other. The 'ischyromyid' rodents are resolved on the rodent stem (*Paramys* and *Reithroparamys*) and one genus, *Ischyromys*, is placed within Crown Rodentia on the stem of Sciuromorpha. The theridomyid rodents are also paraphyletic, *Paradelomys* and *Tarnomys* on the stem of Sciuromorpha and one genus, *Protechimys*, is nested within Myomorpha. The ctenodactyloid rodent *Tataromys* is placed on the Ctenohystrica-Myomorpha stem. *Tataromys* is currently interpreted as a basal member of the Ctenodactyloidea radiation as it possesses an enlarged

infraorbital foramen, broad frontal bone, multiserial incisor enamel and a sciurognathous jaw (Wang, 1997). The placement of *Tataromys* in this topology would imply that Myomorpha is possibly descended from ancestors with ctenodactyloid-like features. Within Ctenohystrica, the Asian hystricognathous fossil *Tsaganomys* is nested within the African group of 'old world hystricognaths', often referred to as 'Phiomorpha' or 'phiomorphs'. The fossil *Gaudeamus*, which has been called the first African caviomorph (Coster et al., 2010), is resolved at the base of the Caviomorpha stem, along with *Paraphiomys.* The latter has been described as a stem phiomorph in other work (Marivaux and Boivin, 2019), although that is not the case here. Finally, Myomorpha is split into two monophyletic clades, a Myodonta + Anomaluromorpha clade and Castorimorpha. *Elymys* is positioned on the stem of the former clade, implying that Emry's (2007) largely correct diagnosis of *Elymys* as the earliest myodont should be extended to the earliest non-castorimorph member of Myomorpha.

The node-dated phylogeny based on my UCE dataset has a number of topological and timing differences from the morph+coding tree described above. Firstly, the UCE topology suggests a much earlier split between Lagomorpha and Rodentia, ca. 120 Ma in the mid Lower Cretaceous (Figure 19). This timing is reminiscent of that found by Kumar and Hedges (1998), but does not reflect more recent estimates (Foley et al. 2016). Secondly, it supports a Ctenohystrica + Sciuromorpha clade to the exclusion of Myomorpha. However, the two topologies offer similar timings for the division into the three main rodent clades. The UCE-based tree implies that Myomorpha split from the Ctenohystrica-Sciuromorpha clade ca. 63 Ma, a couple of millions of years later than timings suggested by Swanson et al. (2019), then soon afterwards, Ctenohystrica split from Sciuromorpha. In general, the branch lengths of many basal relationships in the UCE phylogeny are incredibly small. Figure 22 shows this topology with arbitrary (but legible) branch lengths. Another difference that can be gleaned from the UCE-based tree is that all of the eurymylid taxa bar one form a group on the base of the Lagomorpha stem. As in the morph+coding-based tree, *Sinomylus* is not resolved with the rest of Eurymylidae, but is nested within crown Sciuromorpha alongside Alagomyidae and *Cocomys*, albeit with bootstraps <50. In fact, most of the theridomyids, all of the 'ischyromyids' and many other stem Glires taxa are resolved within the Sciuromorpha crown, which even contains a fossil member of Octodontidae (*Sallamys*; Verzi et al., 2016) and the myomorph *Elymys*. Moving on to Ctenohystrica, *Gaudeamus* and

↓ **Figure 19:** Tip-dated Bayesian phylogeny of Glires based on the morph+coding dataset. The $\delta^{18}O$ data is sourced from Veizer (1999) and Zachos et al. (2001). Generic diversity was estimated from PBDB occurrence records and extrapolated using Alroy's (2018) squares method. Six climatic events are numbered: 1) Paleocene-Eocene Thermal Maximum, 2) Mid-Eocene Climatic Optimum, 3) Oligocene Glaciation, 4) Late Oligocene Warming, 5) Miocene Glaciation and 6) Mid-Miocene Climatic Optimum. A tree with all taxa including non-glires can be found in Appendix 3g.

Figure 19

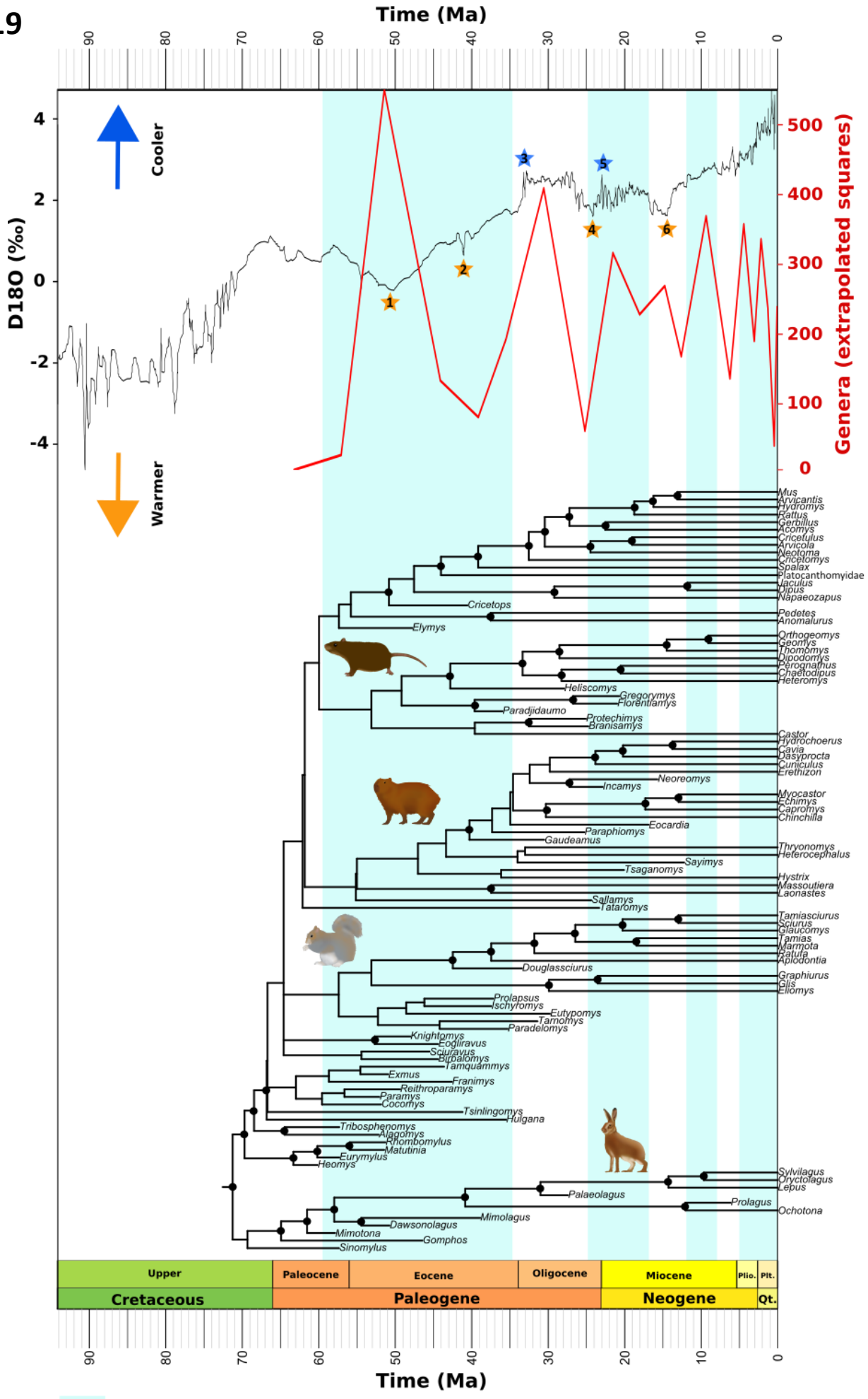
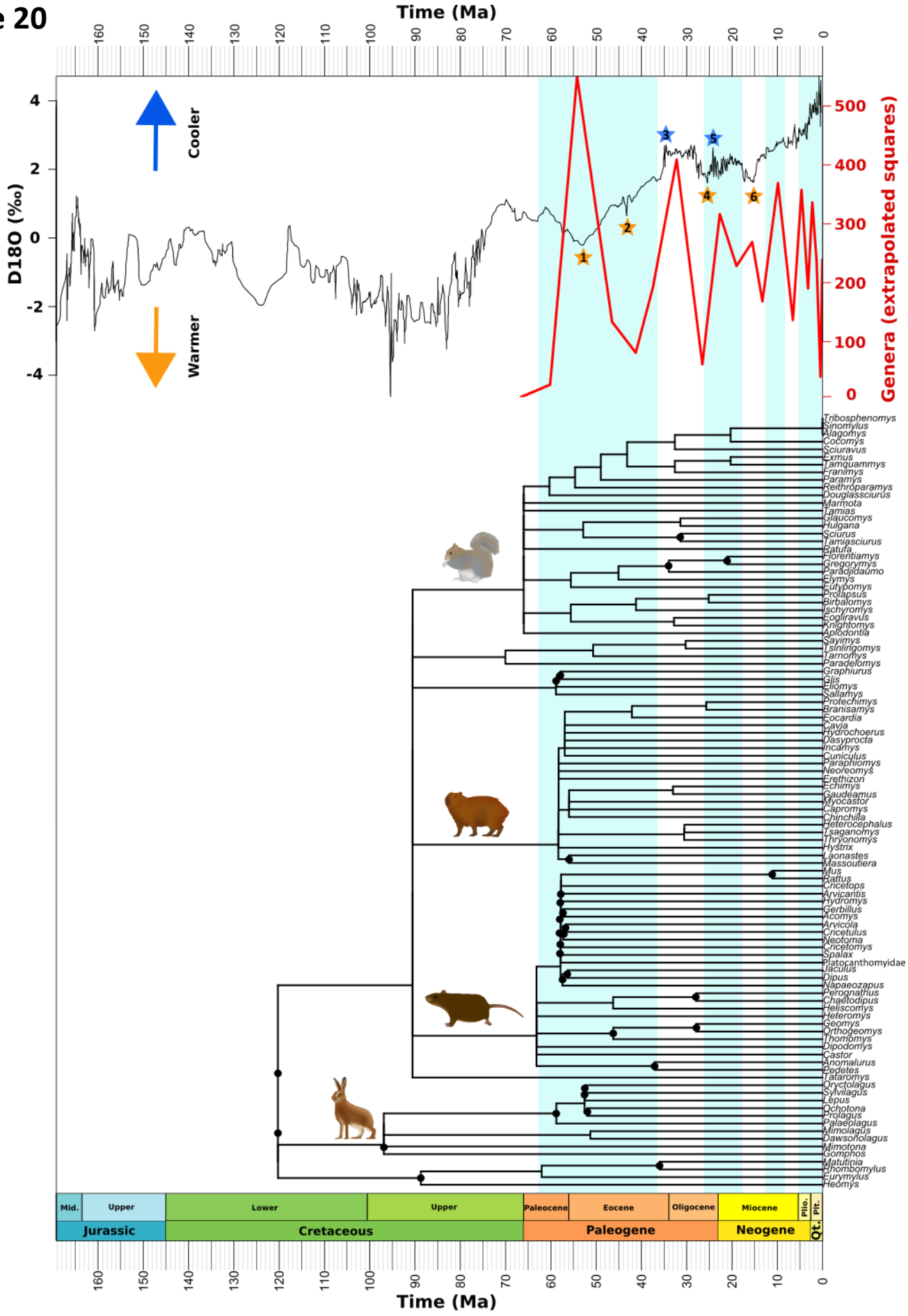

Time (Ma)
90
80
70
60
50
40
30
20
10
0
D18O (‰)
4
2
0
-2
-4
Cooler
Warmer
Genera (extrapolated squares)
500
400
300
200
100
0
1
2
3
4
5
6
Mus
Arvicantis
Hydromys
Rattus
Gerbillus
Acomys
Cricetulus
Arvicola
Neotoma
Cricetomys
Spalax
Platacanthomyidae
Jaculus
Dipus
Napaeozapus
Cricetops
Pedetes
Anomalurus
Elymys
Orthogeomys
Geomys
Thomomys
Dipodomys
Perognathus
Chaetodipus
Heteromys
Heliscomys
Gregorymys
Florentiamys
Paradjidaumo
Protechimys
Branisamys
Castor
Hydrochoerus
Cavia
Dasyprocta
Cuniculus
Erethizon
Neoreomys
Incamys
Myocastor
Echimys
Capromys
Chinchilla
Eocardia
Paraphiomys
Gaudeamus
Thryonomys
Heterocephalus
Sayimys
Tsaganomys
Hystrix
Massoutiera
Laonastes
Sallamys
Tataromys
Tamiasciurus
Sciurus
Glaucomys
Tamias
Marmota
Ratufa
Aplodontia
Douglassciurus
Graphiurus
Glis
Eliomys
Prolapsus
Ischyromys
Eutypomys
Tamonmys
Paradelomys
Knightomys
Eogliravus
Sciuravus
Elpidomys
Tamquammys
Exmus
Franimys
Reithroparamys
Paramys
Cocomys
Tsinlingomys
Hulgana
Tribosphenomys
Alagomys
Rhombomylus
Matutinia
Eurymylus
Heomys
Sylvilagus
Oryctolagus
Lepus
Palaeolagus
Prolagus
Ochotona
Mimolagus
Dawsonolagus
Mimotona
Gomphos
Sinomylus
Upper
Paleocene
Eocene
Oligocene
Miocene
Plio.
Plt.
Cretaceous
Paleogene
Neogene
Qt.
Time (Ma)
Major uplift of the Tibetan Plateau
Warming event
Cooling event
≥ 50% posterior probability

Figure 20



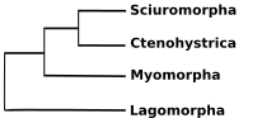
● bootstraps ≥ 50

Major uplift of the Tibetan Plateau

★ Warming event

★ Cooling event

Simplified cladogram of Glires to show clades obscured by short branch lengths



↑ **Figure 20:** Maximum Likelihood phylogeny of Glires based on the UCEs dataset and dated using node calibrations. The $\delta^{18}O$ data is sourced from Veizer (1999) and Zachos et al. (2001) with a rolling 5-point mean for readability. Generic diversity was estimated from PBDB occurrence records and extrapolated using Alroy's (2018) squares method. Climatic events are labelled as in Figure 19 above. A tree with all taxa including non-glires can be found in Appendix 3h.

Protophiomys are both nested within crown Caviomorpha, and *Tsaganomys* remains within the group of old world hystricognaths. The Asian ctenodactyloid fossil *Tataromys* however, roots the whole of Rodentia in the UCE-based tree. This result reflects the position of *Tataromys* found in Asher et al's (2019) phylogeny of Glires. This placement of *Tataromys* would imply a ctenodactyloid-like common ancestor of Rodentia, as opposed to the more traditional anatomically protrogomorphous ancestors implied by the morph+coding phylogeny.

Biogeographic analyses

The morph+coding dataset and the UCEs dataset produce different topologies with different predictions of rodent ancestral ranges. Biogeographic analysis of ancestral ranges plotted onto the morph+coding tree supports an unambiguous Asian origin of Glires, Rodentia and Lagomorpha (Figure 21). However, the predicted ancestral source for Glires is less well-resolved with UCEs. In this case, the common ancestor of Glires would have inhabited Asia, as well as Europe, Africa and North America (Figure 2). Various dispersal routes (Figure 23) linked these continents throughout the Cretaceous until the Late Eocene, although not contemporaneously. The common ancestor of Rodentia is also implied to have had a large range across Asia, Africa and Europe, although not North America. The poorly supported placement (<50 BS) of ancient fossil taxa that are highly nested within crown groups (e.g. Asian fossil taxa within Sciuromorpha), likely contributed to less resolved ancestral ranges predicted for basal nodes in the UCE-based tree. The short branch lengths could also have resulted in these less resolved results, as the rate parameters may have struggled to reach their maximum likelihood values in some instances. With that being said, lack of resolution means that the biogeographic results are imprecise, but not necessarily incorrect.

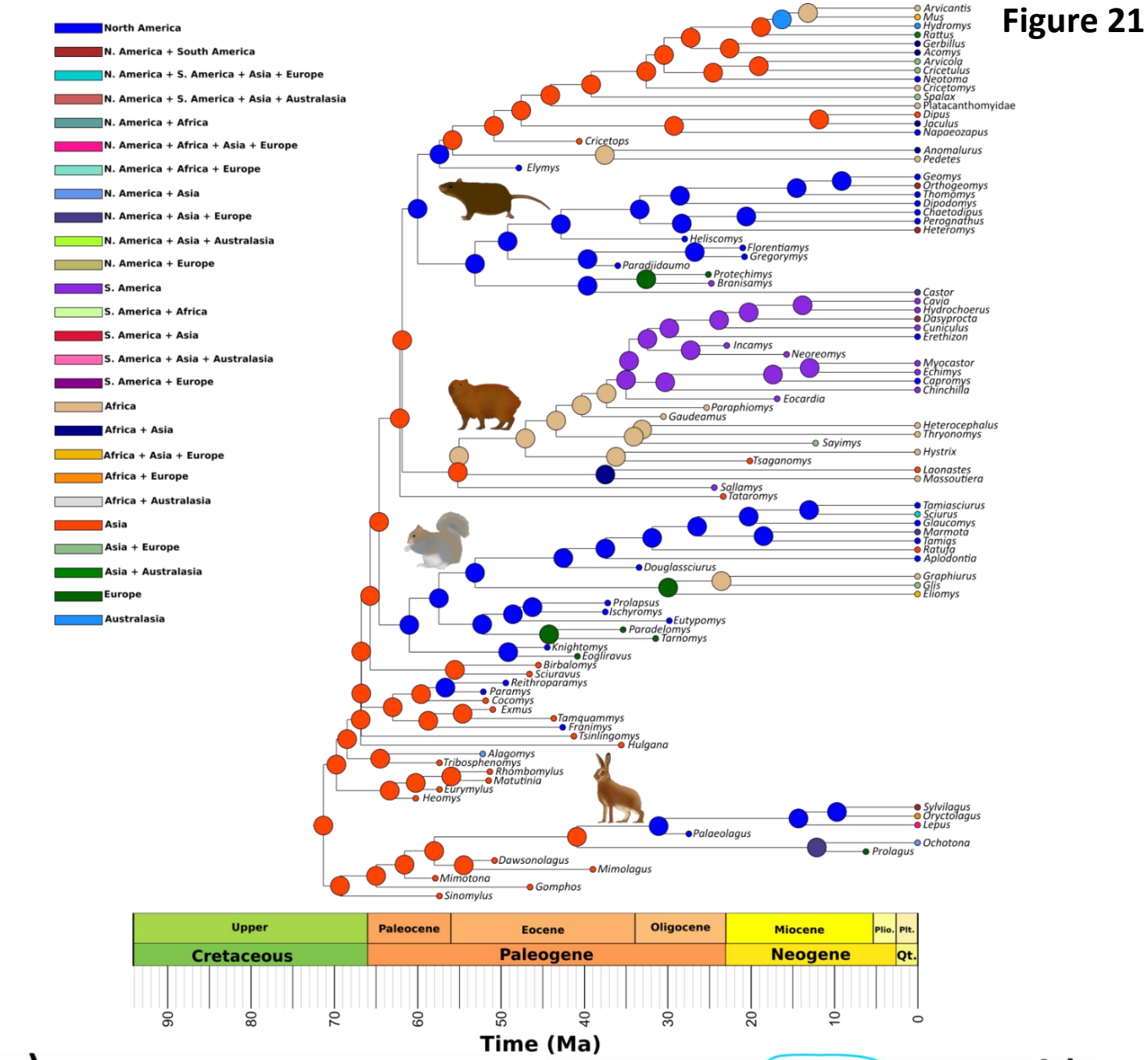
The estimated ancestral ranges of the three living clades differ markedly between the two trees. For example, the ancestral range of the common ancestor of Sciuromorpha is limited to North America in the morph+coding analysis, but due to topological differences in the UCE-based analysis, the ancestral range is almost anywhere other than North America (Africa + Asia + Europe; Figure 22). For the UCE-based tree, this implies that sciuromorphs such as Gliridae diversified across Africa and Eurasia, as well as Theridomyidae in Eurasia, before immigrating to North America in the Upper Cretaceous (Figure 22). The most likely direction of travel at this time, connecting Europe to North America, was the De Geer route (Brikiatis, 2014; Figure 23). According to the UCE-based analysis, the ancestral range of

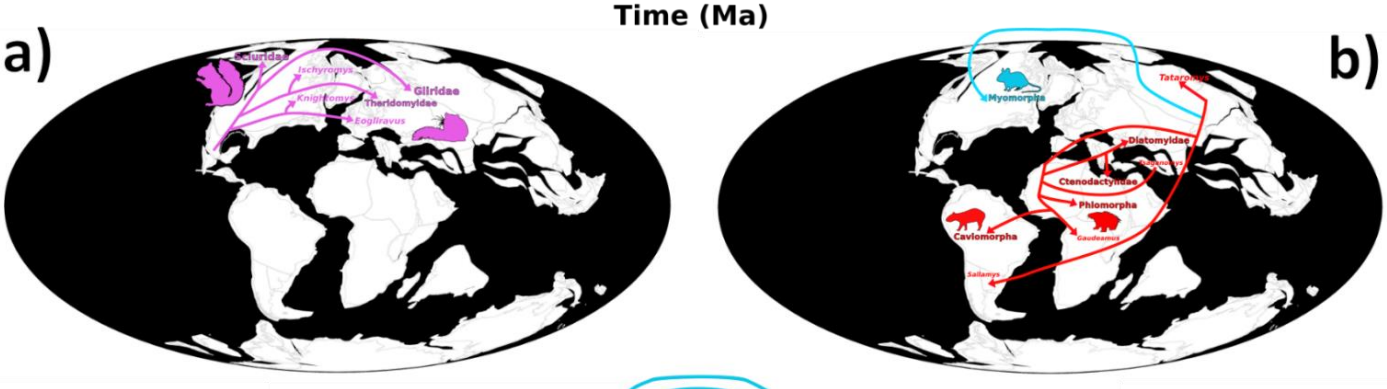
Sciuridae and its stem groups remained across Asia and North America until the Late Eocene, after which, many fossil taxa including the algomyids travelled to Asia. This immigration to Asia would likely have been facilitated by the Bering Land Bridge, which was intermittently passable throughout the Eocene (Brikiatis, 2014). The biogeographic history of Sciuromorpha that is implied by the morph+coding analysis is much simpler. Stem Sciuromorpha taxa make two separate journeys from North America to Europe during the Middle Eocene, first by *Eogliravus* and then by Theridomyidae (Figure 21). The Thulean route, also known as the 'North Atlantic Land Bridge' was closing by the end of the early Eocene so it is possible that these animals passed through Asia first. The morph+coding analysis suggests a European range for the Early Oligocene ancestor of Gliridae, with subsequent dispersals to Africa and Asia. Again, the Bering Land Bridge connected North America and the western continents during this time, and so the exit from North America was likely to have occurred by crossing it.

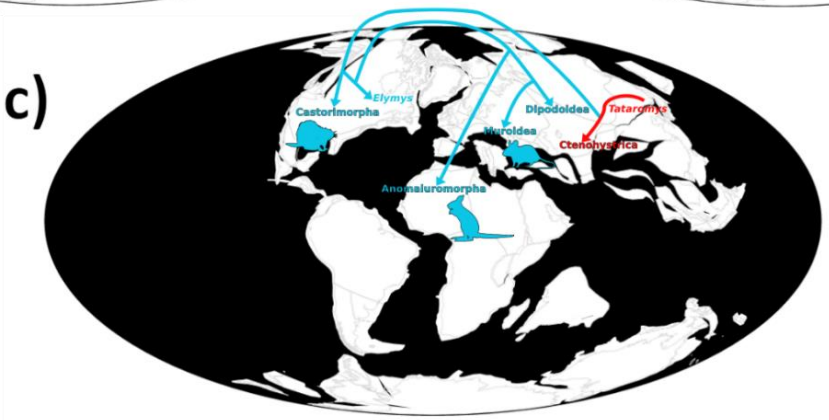
As regards to Myomorpha, the morph+coding analysis suggests that this clade first diversified in North America (Figure 21). However, the Asian origin of the Ctenohystrica + Myomorpha clade implies that at some point, an Asian ancestor of Myomorpha must have crossed the Bering Land Bridge, although the topology indicates that none of the fossil taxa in this study are that ancestor. From North America across the Bering Land Bridge, an early dispersal to Asia occurs during the latest Paleocene-Early Eocene, resulting in the diversification of Anomaluromorpha and Myodonta. The version of events according to the UCE dataset analysis provides an opposite series of events. In this case, Myomorpha inhabited a wide range across Europe, Africa and Asia for the most part of their history, with this range contracting to Asia + Africa in Anomaluromorpha. The predominantly North American Castorimorpha left Asia early on via either the De Geer route or Bering Land Bridge, which were both crossable during the early Paleocene (Figure 23).

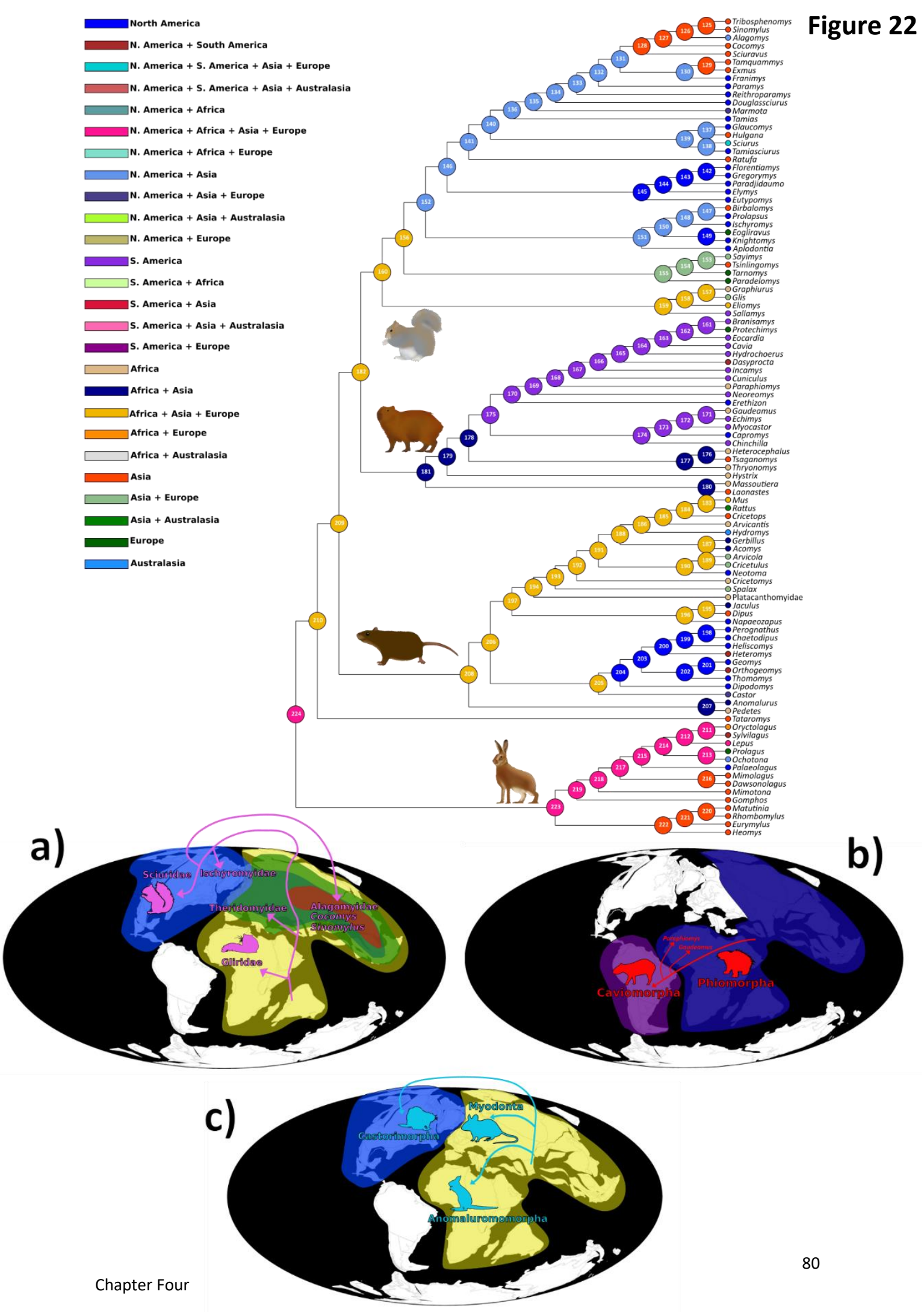
Finally, the series of dispersals within Ctenohystrica also differ between the two total-evidence phylogenies. In the UCE analysis, the common ancestor to all Ctenohystrica was spread across Africa and Asia. In the morph+coding analysis, the common ancestor of crown group Ctenohystrica was restricted to Africa, but if total group Ctenohystrica is considered, then Asia is resolved as the point of origin. According to the UCE analysis, the common ancestor of Hystricognathi and Ctenodactylomorphi was Afro-Asian and the dispersal of caviomorphs from this region to South America took place ca. 59 million years ago in the Late Paleocene, roughly coinciding with the

↓ **Figure 21:** Most likely estimated ancestral ranges for Glires based on the tip-dated Bayesian morph+coding phylogeny and a DIVALIKE+J model of dispersal. Simplified hypotheses of Glires dispersal have been superimposed onto Paleocene maps for all three crown groups: a) Sciuromorpha, b) Ctenohystrica and c) Myomorpha. Maps sourced from PBDB. A tree with all taxa including non-glires can be found in Appendix 3i.


North America
N. America + South America
N. America + S. America + Asia + Europe
N. America + S. America + Asia + Australasia
N. America + Africa
N. America + Africa + Asia + Europe
N. America + Africa + Europe
N. America + Asia
N. America + Asia + Europe
N. America + Asia + Australasia
N. America + Europe
S. America
S. America + Africa
S. America + Asia
S. America + Asia + Australasia
S. America + Europe
Africa
Africa + Asia
Africa + Asia + Europe
Africa + Europe
Africa + Australasia
Asia
Asia + Europe
Asia + Australasia
Europe
Australasia
Arvicantis
Mus
Hydromys
Rattus
Gerbillus
Acomys
Arvicola
Cricetulus
Neotoma
Cricetomys
Spalax
Platacanthomyidae
Dipus
Jaculus
Napaeozapus
Cricetops
Anomalurus
Pedetes
Elymys
Geomys
Orthogeomys
Thomomys
Dipodomys
Chaetodipus
Perognathus
Heteromys
Heliscomys
Florentiamys
Gregorymys
Paradjidaumo
Protechimys
Branisamys
Castor
Cavia
Hydrochoerus
Dasyprocta
Cuniculus
Erethizon
Incamys
Neoreomys
Myocastor
Echimys
Capromys
Chinchilla
Eocardia
Paraphiomys
Gaudeamus
Heterocephalus
Thryonomys
Sayimys
Hystrix
Tsaganomys
Laonastes
Massoutiera
Sallamys
Tataromys
Tamiasciurus
Sciurus
Glaucomys
Marmota
Tamias
Ratufa
Aplodontia
Douglassciurus
Graphiurus
Glis
Eliomys
Prolapsus
Ischyromys
Eutypomys
Paradelomys
Tarnomys
Knightomys
Eogliravus
Birbalomys
Sciuravus
Reithroparamys
Paramys
Cocomys
Exmus
Tamquammys
Franimys
Tsinlingomys
Hulgana
Alagomys
Tribosphenomys
Rhombomylus
Matutinia
Eurymylus
Heomys
Sylvilagus
Oryctolagus
Lepus
Palaeolagus
Ochotona
Prolagus
Dawsonolagus
Mimolagus
Mimotona
Gomphos
Sinomylus
Upper
Paleocene
Eocene
Oligocene
Miocene
Plio.
Plt.
Cretaceous
Paleogene
Neogene
Qt.
90
80
70
60
50
40
30
20
10
0
Time (Ma)


a)
Sciuridae
Ischyromys
Knightomys
Theridomyidae
Gliridae
Eogliravus
b)
Myomorpha
Tataromys
Diatomyidae
Ctenodactylidae
Phiomorpha
Caviomorpha
Gaudeamus
Sallamys


c)
Elymys
Castorimorpha
Dipodoidea
Tatarmys
Myroidea
Ctenohystrica
Anomaluromorpha


North America
N. America + South America
N. America + S. America + Asia + Europe
N. America + S. America + Asia + Australasia
N. America + Africa
N. America + Africa + Asia + Europe
N. America + Africa + Europe
N. America + Asia
N. America + Asia + Europe
N. America + Asia + Australasia
N. America + Europe
S. America
S. America + Africa
S. America + Asia
S. America + Asia + Australasia
S. America + Europe
Africa
Africa + Asia
Africa + Asia + Europe
Africa + Europe
Africa + Australasia
Asia
Asia + Europe
Asia + Australasia
Europe
Australasia
Tribosphenomys
Sinomylus
Alagomys
Cocomys
Sciuravus
Tamquammys
Exmus
Franimys
Paramys
Reithroparamys
Douglassciurus
Marmota
Tamias
Glaucomys
Hulgana
Sciurus
Tamiasciurus
Ratufa
Florentiamys
Gregorymys
Paradjidaumo
Elymys
Eutypomys
Birbalomys
Prolapsus
Ischyromys
Eogliravus
Knightomys
Aplodontia
Sayimys
Tsinlingomys
Tarnomys
Paradelomys
Graphiurus
Glis
Eliomys
Sallamys
Branisamys
Protechimys
Eocardia
Cavia
Hydrochoerus
Dasyprocta
Incamys
Cuniculus
Paraphiomys
Neoreomys
Erethizon
Gaudeamus
Echimys
Myocastor
Capromys
Chinchilla
Heterocephalus
Tsaganomys
Thryonomys
Hystrix
Massoutiera
Laonastes
Mus
Rattus
Cricetops
Arvicantis
Hydromys
Gerbillus
Acomys
Arvicola
Cricetulus
Neotoma
Cricetomys
Spalax
Platacanthomyidae
Jaculus
Dipus
Napaeozapus
Perognathus
Chaetodipus
Heliscomys
Heteromys
Geomys
Orthogeomys
Thomomys
Dipodomys
Castor
Anomalurus
Pedetes
Tataromys
Oryctolagus
Sylvilagus
Lepus
Prolagus
Ochotona
Palaeolagus
Mimolagus
Dawsonolagus
Mimotona
Gomphos
Matutinia
Rhombomylus
Eurymylus
Heomys
a)
Sciuridae
Ischyromyidae
Theridomyidae
Alagomyidae
Cocomys
Sinomylus
Gliridae
b)
Paraphiomys
Gaudeamus
Caviomorpha
Phiomorpha
c)
Myodonta
Castorimorpha
Anomaluromorpha

↑ **Figure 22:** Most likely ancestral ranges estimated for Glires based on the node-dated Maximum Likelihood UCEs phylogeny and a BAYAREALIKE+J model of dispersal. Simplified hypotheses of Glires dispersal have been superimposed onto Late Cretaceous maps for all 3 crown groups: a) Sciuromorpha, b) Ctenohystrica and c) Myomorpha. Maps sourced from PBDB. A tree with all taxa including non-glires can be found in Appendix 3j.

Paleocene-Eocene Thermal Maximum (Zachos et al, 2008). A reversal back to Africa took place twice, once for *Protophiomys* (almost immediately after arrival in South America) and secondly for *Gaudeamus* (ca. 33 Ma). The morph+coding biogeographic analyses suggest that the common ancestor of Hystricognathi and Ctenodactylomorphi was African. Old world hystricognaths then diversified across Africa, as did the first caviomorphs (*Gaudeamus*), before dispersal to South America took place ca. 35 Ma during the Late Eocene. This roughly coincides with the Eocene-Oligocene global cooling and drying period (Zachos et al., 2008).

Diversification and the uplift of the Tibetan Plateau

Several peaks in Glires diversification appear to coincide with increased tectonic activity and climatic events. Once again, whether or not the phylogenetic analyses incorporate UCEs affects how the results are interpreted. According to the morph+coding topology, the split between Sciuromorpha and the Ctenohystrica + Myomorpha clade occurs 6 Ma before the first signs of elevation caused by the approaching Indian plate (Figure 19). However, the first peak in Glires diversity, according to my extrapolated squares estimates, coincides with two very important dates: the collision of the Indian plate with Asia and the Paleocene-Eocene Thermal Maximum. The topology implies that many rodent groups first appeared during this period of tectonic upheaval in Asia and warm temperatures. These groups include predominantly Asian Myodonta, Afro-Asian Anomaluromorpha and some American groups too, e.g. Ischyromyidae and Castorimorpha. The UCE-based topology also suggests that the crown rodent groups split from each other before the first signs of uplift in Southeast Asia. The short branches on this tree indicate an explosive radiation, which resulted in almost all groups within Myomorpha and Ctenohystrica appearing during the first period of uplift in the Tibetan Plateau. The predominantly North American Sciuromorpha however, start to rapidly diversify before this period of uplift in the much cooler latest Cretaceous-earliest Paleocene climate. The second spike in Glires diversity according to the extrapolated squares estimates, occurs during a cold-snap in the Oligocene, which heralded the onset of Antarctic ice sheets (Zachos et al., 2001). Looking at the morph+coding phylogeny, this peak of diversity occurs at the same time as the appearance of the first dipodids and other myodont taxa of Asian descent, as well as Afro-Asian glirids, North American geomyoids, Sciuridae and South American Caviomorpha. The next period of significant uplift in the Tibetan Plateau region co-occurs with two major climatic events, firstly the Late Oligocene warming and then the Miocene glacial expansion (Zachos et al., 2001). It is this second, much cooler, event that correlates

with the third peak in rodent diversity (Figure 19). The morph+coding topology describes the appearance of Asian cricetid and muroid groups during this time, as well as South American Cavioidea, North American Perognathinae and sciurids. There are a further five peaks in rodent diversity leading up to the present day, the three largest of which are estimated to occur during periods of significant Tibetan Plateau uplift.

It is important to note that although many of the peaks in Glires diversification appear to correlate with increased tectonic activity and climatic events by eye, it is necessary to test such trends using statistical inference. The Spearman's test finds a moderate positive correlation between peaks in Glires diversity and periods of intense uplift of the Tibetan Plateau (rho = 0.301). However, the significance of the test is low (p = 0.1). For the oxygen isotope data, there is a small positive correlation between generic diversity and the percentage of $\delta^{18}O$ (rho = 0.222), which indicates that cooler global temperatures are associated with higher generic diversity. The significance of this correlation is also low (p = 0.17). It is worth noting however, that the sample size of my dataset is quite small for such a test. The extrapolated squares method of estimating diversity requires binned data and in this case there is a diversity estimate for each of the 21 geological stages that Glires have been present in. Using smaller time bins could increase the number of data points representing generic diversity, but this might have the undesired effect of inflating the extrapolated squares estimate, due to increased incidence of singletons. In addition, the resolution of the Tibetan Plateau data is unfortunately limited. The lack of a continuous elevation record and non-contemporaneous activity across what is a large landmass, means that the tectonic activity of the Tibetan Plateau is still best represented by merging large chunks of overlapping activity. A post-hoc power analysis can determine the number of observations required to detect an effect of a given size with a specific level of confidence. Taking my Spearman's test results as an example, a correlation coefficient of 0.301 between generic diversity and Tibetan Plateau uplift, and an effect size of 0.8 (implying a conventional four-to-one trade off between Type II and Type I error), post-hoc power analysis suggests that an estimated 66 data points would be needed to find a significant effect. For a correlation coefficient of 0.222 between generic diversity and percentage of $\delta^{18}O$, and a power of 0.8, 122 data points would be needed to detect a significant effect. In both cases, post-hoc power analysis indicates that more data are required. Given the limitations of my data and uncertainty regarding the strength of the observed trends, any remarks based on the correlations between diversity, global temperatures and tectonic activity must be approached tentatively.

Discussion

Ultraconserved elements and the total-evidence approach

A common theme throughout this project has been discovering what effect analysing UCEs has on phylogenetic analyses that include fossil data. Combining these small pieces of non-coding DNA with morphological data is an understudied area, but one of the few studies to include fossil data and UCEs was conducted recently by Chen et al. in 2019. They found that the positions of fossils differed greatly between their total-evidence analyses, although they were unsure of the cause. In one analysis, morphology, protein coding DNA and UCEs were allowed to interact, and in another, a morphology + protein coding DNA matrix was constrained by a backbone based solely on UCE data. In this latter analysis, the phylogenetic signal from UCEs was present, but the UCE data itself was not allowed to mix with the other partitions, thus ignoring any hidden signal that might have been common to all of the data partitions. Hidden signal (Gatesy et al., 1999) shared between all three of my data partitions could be the cause of competing hypotheses of fossil placement in my two analyses. Another possible explanation of course, and one which leads us into more difficult territory, is the idea that conflicting signal from UCEs over-influence phylogeny. The sheer volume of UCE data that can be incorporated into an analysis could conceivably tip the balance of phylogenetic signal overwhelmingly in their favour. The UCE alignment is 914,885bp long, which is considerably larger than the protein coding DNA alignment of 15,407bp, and much larger still than my morphological partition, which is 219 characters long. For fossil taxa, this represents an enormous amount of missing data. It seems plausible that the poorly supported and spurious placement of some of my fossils in the UCE-based topology, such as *Birbalomys* in Sciuromorpha and *Protechimys* in Ctenohystrica, could be because of this data gap. Non-random distribution of missing data can be problematic for Maximum Likelihood methods and produce topological artefacts that Bayesian inference is more robust to (Simmons, 2011; Simmons, 2014). However, there are also plenty of fossils which do achieve the phylogenetic placements that we would expect: for example, *Gaudeamus*, *Protophiomys* and *Tsaganomys* in Ctenohystrica, *Heliscomys* and *Cricetops* in Myomorpha and *Ischyromys* and *Douglassciurus* in Sciuromorpha. If morphological signal had little to contribute in my analyses, we would expect fossils to be scattered randomly across the UCE-based tree and that is not the case here. In addition, of the living taxa in my matrix lacking UCEs (and therefore with 98% missing data), the majority appear in the same positions (Figure 20) as in my morph+coding-based tree (Figure 19), such as *Arvicola* in Myodonta, *Erethizon* in Hystricognathi and *Glaucomys* in Sciuromorpha. Whether or not missing data have a negative effect on topology is still debated and is discussed in detail in Chapter Two and Chapter Three. There are studies such as Wiens and Moen (2008) that suggest that even when a fossil taxon is 95% incomplete, it can still achieve accurate phylogenetic placement in topologies based on

morphology and molecular data. However, there may be an upper limit to the ability of a given dataset to handle large amounts of missing data. In Scarpetta's (2020) investigation of combined UCEs and morphology data, he found that morphological signal was dwarfed by the signal contained within UCEs. This effect was lessened when he increased the weight of morphological characters relative to the UCEs. However, morphological characters are not equivalent to molecular base pairs and we should consider carefully if artificially altering their influence on phylogeny is in fact desirable. Some researchers choose to subsample or reduce the size of their genomic datasets in order to avoid the need for such weighting (Tucker et al., 2017). However, excluding available genomic data from matrices that include morphology defeats the arguments of why you might choose to combine them in the first place. Total-evidence approaches have the ability to amplify shared signal that otherwise might not come to the fore, but this opportunity is missed when the data are not there in the first place. So far, there is no obvious solution on how best to combine fossil and UCE data. Rather than concentrating on how to balance our UCE data with morphology, I would focus future research on improving the options for analysing large datasets in general. My analysis of the UCE dataset was limited by computational power and prevented me (and many others) from adopting a Bayesian inference approach as well as tip-dating and more complex models of evolution. Yet for total-evidence analyses, a Bayesian algorithm that takes into account prior probabilities inferred from real fossil data, seems a more intuitive way to study UCEs and fossil material together. For these reasons, my morph+coding topology, based on a Bayesian tip-dating approach, may be preferable to my UCE-based topology, as Maximum Likelihood and penalized likelihood dating methods may be suboptimal for this type of data. Nevertheless, my results indicate that informative results can be achieved by combining genomic and fossil data, although developing technology to allow more complex analyses of mixed data will greatly benefit this kind of research.

The rodent root

Glires have a good fossil record and good molecular sampling, but even though we have learnt much about their evolutionary history throughout the years, there are still parts of their past that require clarity. As evidenced by my own results, there are questions surrounding the basal relationships and ancient biogeography of Glires that are not easily answered. The relationships between the three living rodent groups (Sciuromorpha, Ctenohystrica and Myomorpha) are a prime example of this. In my phylogenetic analysis of morphology and protein coding loci, Sciuromorpha is resolved at the root of Rodentia and a sister relationship is found between Ctenohystrica and Myomorpha (Figure 19). There are many studies which share this result, including Douzery et al. (2003), Churakov et al. (2010), Fabre et al. (2012) and Asher et al. (2019). Given that many stem Sciuromorpha taxa and stem Glires taxa are anatomically protrogomorphous, this relationship is historically the one most cited in the

literature. In contrast, when ultraconserved elements are added to the matrix, Myomorpha comprises the root (Figure 20). This combination has also been observed in studies by Huchon et al. (2000), DeBry (2003), Montgelard et al. (2008), Meredith et al. (2011) and Swanson et al. (2019). A third possible relationship, where Ctenohystrica roots Rodentia, has also been considered by others (Huchon et al., 2007). With so many conflicting topologies, the root of Rodentia is perhaps best imagined as an unresolved trichotomy for the time being. While this may seem conservative, it is clear from my results that the type of data being analysed, and perhaps the volume too, has a big effect on topological relationships. Another fact to consider is that my phylogenetic analyses suggest that Rodentia split into the three clades very quickly, in what was a geologically explosive radiation. Both my morph+coding and UCEs-based trees describe how rodents were already remarkably diversified by the end of the Paleocene. When discussing the trichotomy at the root of Rodentia, Blanga-Kanfi et al. (2009) postulated how such a rapid radiation could conceivably result in incomplete lineage sorting. When cladogenetic events occur in quick succession (or even simultaneously by different isolated populations), loci may fail to coalesce during speciation, and so teasing apart the order of these events becomes tricky. Nevertheless, my observation that Glires diversified during the Paleocene provides further evidence for the Cenozoic heralding the 'Age of Mammals', as indeed many other mammal groups appear during this period of intense global warming (Beard, 1998).

The placement of *Tataromys*

Several key fossil taxa occupy different positions on my two topologies and imply contrasting hypotheses of rodent evolution. The position of *Tataromys*, an anatomically sciurognathous and hystricomorphous rodent, is of interest as it is considered to be a basal member of Ctenohystrica. My morph+coding-based tree places *Tataromys* on the Ctenohystrica + Myomorpha root (Figure 19). *Tataromys* is a well-recognised ctenodactlyoid fossil (Wang, 1997; Oliver et al., 2017) and its placement here could imply that both Ctenohystrica and Myomorpha are descended from an ancestor with ctenodactyloid-like features. This is not the first time such a connection has been made. As early as 1899, Tullberg realised the similarity in jaw structure between Myomorpha and Ctenodactylomorphi, and as a consequence he allocated the latter to the Myomorpha. For similar reasons, Simpson (1945) included certain ctenodactlyids within Myomorpha under *incertae sedis*. Later on, Vianey-Liaud (1985) suggested that dipodid and cricetid myomorphs in particular could have evolved from Asian ctenodactyloid ancestors. However, my UCE-based phylogeny places *Tataromys* at the root of Rodentia (Figure 20). Interestingly, this result is shared by Asher et al. (2019) who did not use UCEs in their matrix but a very similar 16k-DNA alignment (expanded here with the inclusion of *Cricetulus*) plus the same 219 morph characters used here (expanded with 21 fossil taxa). The idea of ctenodactyloid fossils representing the most basal rodents or stem group rodents is perhaps not

too far-fetched given the antiquity of ctenodactyloid fossils in Asia. *Cocomys* for example, is considered to be the earliest fossil with ctenodactyloid-like features (Li et al., 1989) and is found on the stem of Rodentia in my morph+coding tree (Figure 19). In fact, several authors have proposed the possibility that Ctenodactyloidea could be either the earliest rodent group (Korth, 1984; Dawson et al., 1984) or a stem group of Rodentia (Hartenberger, 1985), which presumably originated in the Eocene of Asia. The different phylogenetic affinities of *Tataromys* highlight the unexpected effects of analysing different types of data together, and the consequences for our ideas of rodent evolution.

Biogeography

The Asian origin of Glires has never been seriously disputed given their strong representation in the Paleocene deposits of Asia (Li, 1977; Li et al., 2016). Yet the presence of ischyromyid and alagomyid fossil taxa in Paleocene of North America has prompted discussion of a possible North American origin of Glires in the past (Sloan, 1969; Wood, 1977). My morph+coding ancestral range estimation analysis unequivocally supports an Asian origin for Glires, Rodentia and Lagomorpha (Figure 21). This adds to the growing body of evidence that many mammal groups, such as Primates, Perissodactyls and Artiodactyls, originated on the Asian continent (Beard, 1998). If the Cenozoic era is considered the 'Age of Mammals', then Asia would surely be the 'cradle' from which they emerged, or as Beard put it, "The Garden of Eden" (2002, p. 2029). As the appearance of many mammal groups coincided with the Paleocene-Eocene Thermal Maximum, it is thought that this period of global warming was the driver behind the reorganization of mammals (Bowen et al., 2002). My UCE-based analysis provides less resolution as regards to the geographic origins of Glires (Figure 22). In this case, the estimated ancestral range leaves open the possibility that glires were already spread across the Northern Hemisphere before they diversified. There is no fossil evidence to suggest that the ancestors of Glires occupied any continent other than Asia, although as with any other mammal group, new fossil discoveries could change this at any time.

The importance of new fossil and living species finds is exemplified by the rodent group Ctenohystrica. Until the recent discovery (to the academic community) of *Laonastes aenigmamus* in 2006, living members of Ctenohystrica were considered, with the exception of Hystricidae, to be exclusively African and South American. However, we now know that members of Ctenohystrica did occupy Asia in much greater numbers, first through ctenodactyloids (e.g. *Tataromys* and *Cocomys*) and later on by members of Hystricognathi (e.g. *Tsaganomys*). Curiously, the first hystricognaths do not appear in Asia until many years after their first appearance in Africa and South America, although an Asian origin is heavily implied in the literature (Antoine et al., 2012; Marivaux et al., 2014; Boivin, 2017; Boivin et al., 2019). My results confirm that the common ancestor of all Ctenohystrica was indeed Asian or had perhaps occupied both Asia and Africa before diversifying into the groups we see today (Figures 21

and 22). However, while my UCE-based analysis suggests that Hystricognathi split from Ctenodactylomorphi in the Afro-Asian region, my morph+coding results imply that this occurred specifically in Africa. This latter result would imply that the ctenodactyloid family Diatomyidae, and Oligocene hystricognaths such as *Tsaganomys,* made the journey back to Asia on separate occasions. It also implies that there could be ctenodactyloid ancestors of Hystricognathi in Africa, as of yet undiscovered. These results are only in partial agreement with Marivaux and Boivin's (2019) study on the origin of hystricognathous rodents. Marivaux and Boivin described several tentative hypotheses of dispersal and although they included reversals from Africa back to Asia, they found no evidence for Hystricognathi originating in Africa. My total-evidence analyses also struggle to provide a consensus on the origin of caviomorph rodents. The morph+coding results (Figure 21) imply that caviomorphs originated in Africa before immigrating to South America ca. 35 Ma, coinciding with the Eocene-Oligocene global cooling and drying period (Zachos et al., 2008). This finding mirrors Poux et al's (2006) timings for caviomorph dispersal, which is thought to have been aided by rafting or stepping stone islands (Wyss et al., 1993; Flynn and Wyss, 1998; Houle, 1999). It has been hypothesised that the hypsodont teeth of caviomorphs were an adaption to the expansion of grasslands and open woodland during the cooling period (Vucetich et al. 1999; Poux et al., 2006). Hypsodont teeth would certainly allow rodents who possessed them to exploit these new niches and diversify. My UCE-based analysis however, predicts a much earlier dispersal to South America ca. 59 Ma during the time of the Paleocene-Eocene Thermal Maximum (Figure 22). Neither of these figures however, likely reflects the true order of events. The discovery stem caviomorph dental remains from the Yahuarango Formation of Peru imply that caviomorphs were in South America 41 Ma in the Middle Eocene (Antoine et al., 2012). The presence of these teeth and dated phylogenies by Antoine et al. (2012) suggest that caviomorphs actually immigrated to South America during the Mid-Eocene Climatic Optimum, when the climate was warm and wet. However, this timing of events poses a problem of its own. Hystricognath fossils are not known in Africa until the latest Middle Eocene and the oldest African caviomorph fossil (*Gaudeamus*) is only recorded from the early Oligocene. This would imply ghost lineages for both Hystricognathi and Caviomorpha. The only way to reconcile this disparity would be to find new fossil hystricognaths in Africa that could fill the gaps. It is also possible that *Gaudeamus* represents convergent evolution of the caviomorph type of dentition (Coster et al., 2010; Antoine et al., 2012). In this case, the ancestors of *Gaudeamus* could have been the ones to travel to South America, thus potentially reducing the ghost lineage implied by the earliest South American caviomorphs. These tentative hypotheses require further evidence and emphasise the need to continue sampling the fossil record, as incomplete sampling will bias the results of any biogeographic analyses.

With movement of Ctenohystrica westwards from Asia confirmed, the dispersal of Myomorpha and Sciuromorpha can now be examined. Unsurprisingly, both of my total evidence analyses suggest that Myomorpha split from other rodents in Asia (range extended to include Africa + Eurasia in the UCE-based analysis; Figures 21 and 22). However, the origin of the most recent common ancestor of Myomorpha has two possible hypotheses. The ancestral range reconstruction based on my morph+coding data proposes a North American origin of Myomorpha, but my results for UCEs suggest pretty much anywhere other than North America (Africa + Eurasia). An identical result is found for Sciuromorpha as well. These differences of course, have very large consequences for the biogeographic histories of Sciuromorpha and Myomorpha. For Sciuromorpha, one scenario implies the diversification of stem Sciuromorpha in North America (including the 'ischyromyids') before multiple dispersal events to Europe during the Eocene and Oligocene (via the Thulean route and then the Bering Land Bridge, Figures 21 and 23). The other scenario points towards diversification across Eurasia before dispersals to North America during the Late Cretaceous (presumably via the De Geer route, Figures 22 and 23), before a reversal back to Asia in the Late Eocene (via the Bering Land Bridge). The latter series of events clearly imply timings that are not reconcilable with the fossil record. However, the alternative of deep-rooted origins in North America are also controversial. For Myomorpha, the UCE-based analysis proposes an Africa + Eurasia wide diversification with immigration to North America taking place during the early Paleocene (either across the Bering Land Bridge or the De Geer Route, Figures 22 and 23). The morph+coding results go in the opposite direction, with stem myodonts (such as *Elymys*) diversifying in North America and then an early dispersal from North America to Asia during the latest Paleocene-Early Eocene, followed by diversification of stem cricetids and extant myodonta (via the Bering Land Bridge). While a North America origin for any rodent group is generally met with scepticism, the latter sequence of events is remarkably similar to that hypothesised by Emry (2007). He hypothesised that Myomorpha first originated in North America before an *Elymys*-like ancestor travelled to Asia sometime during the Eocene, leading to the diversification of Myodonta in Asia. While my results offer conflicting ideas of sciuromorph and myomorph origins, it is clear that North America did play a large role in the evolution of both of these groups. Rather than being a 'cul de sac' of Asia, my results suggest that North America could have been the centre of origin for the most speciose rodent group. With more certainty, they suggest that dispersal between Asia and North America was frequent.

The Tibetan Plateau

As discussed earlier, a growing body of evidence points towards Asia as the centre of early mammal (including Glires) diversification, and now at least some of that activity could be explained by the formation of the Tibetan Plateau. While my total-evidence analyses both support an Upper Cretaceous

origin of Glires, the first and largest spike in Glires diversity coincides with the collision of the Indian plate with Asia and the Paleocene-Eocene Thermal Maximum (PETM) (Figures 19 and 20).

There are plenty of reasons why climatic and tectonic events might be responsible for changing patterns in biodiversity. The Paleogene is a time period known for substantial climatic and topographic change, going from a hothouse world to an icy landscape in just 15 million years, as well as large changes in sea-level and landmass positions, leading to the formation of new land bridges. The PETM is associated with large mammalian turnovers across Asia, Europe, North Africa and North America. It has been hypothesised that changes in mammalian diversity came about via ecological responses to global warming, as well as through dispersal events across newly connected lands (Hooker 2000). For example, Bighorn Basin in Wyoming has one of the best studied records of terrestrial life during the PETM. Studies have found that the vegetation of Bighorn Basin transitioned from temperate to subtropical flora, while the body size of mammalian fauna underwent 'dwarfing', including the many immigrant taxa that had only just appeared, e.g. primates, artiodactlys and perissodactlys (Clyde and Gingerich, 1988; McInerney and wing, 2011; Record et al., 2012). Some explanations for the changes in body size have drawn upon Bergmann's rule, as mammals living in warmer climates tend to have larger surface area-to-volume ratios, but increased incidence of drought and nutrient availability have also been used to explain changes in mammalian morphology around the PETM (D'Ambrosia et al., 2017). As well as the PETM itself, significant topographic changes were occurring in Asia due to the collision of the Indian plate. The emission of large quantities of CO^2, deriving from the subduction and continental arc volcanism caused by the collision, likely contributed to the PETM and so the two events cannot be easily teased apart (Kent et al., 2003; Chatterjee et al., 2017). The impact of the collision created one of the most tectonically active mountain ranges on Earth, with new ecological opportunities leading to the Great Indo-Eurasian Interchange (Chatterjee et al, 2017). The complex new topography of the region, caused by the orogeny of the Himalayas and the drainage of rivers, is known to have created and removed barriers, leading to altered patterns of vicariance and therefore, biodiversity (Shih et al., 2009; Rüber et al., 2020).

It is certainly possible that harsh conditions and varied topography imposed by the collision, coupled with rising global temperatures, could have opened up new niches and encouraged speciation within Glires. If this is the case, it would support Badgley's (2010) hypothesis that extreme tectonic events can lead to species diversification. In fact, my UCE-based topology suggests that most of the rodent diversity we see today evolved during this time (Figure 20), whereas my morph+coding results highlight early diversification of myodonts and ctenodactyloids in Asia, as well as ischyromyids and castorids in North America (Figure 19). After this peak in diversity, rodents experience a very sharp

decline. However, a further five peaks in Glires diversity (ca. 21, 10, 4.5, 2.5 Ma and present day) correlate with periods of intense uplift in the Tibetan Plateau area.

It is difficult to prove definitively that a particular climatic or tectonic event in the past was directly responsible for the evolutionary history of an animal group. The Spearman's correlation tests carried out in this study do suggest that periods of intense uplift and cooler temperatures may have had a positive effect on glires biodiversity, but lack of statistical power combined with necessarily unprecise timings of the Tibetan Plateau uplift schedule make it difficult to make these claims with certainty. It is therefore sensible to treat remarks on diversification and its potential drivers with tentative caution.

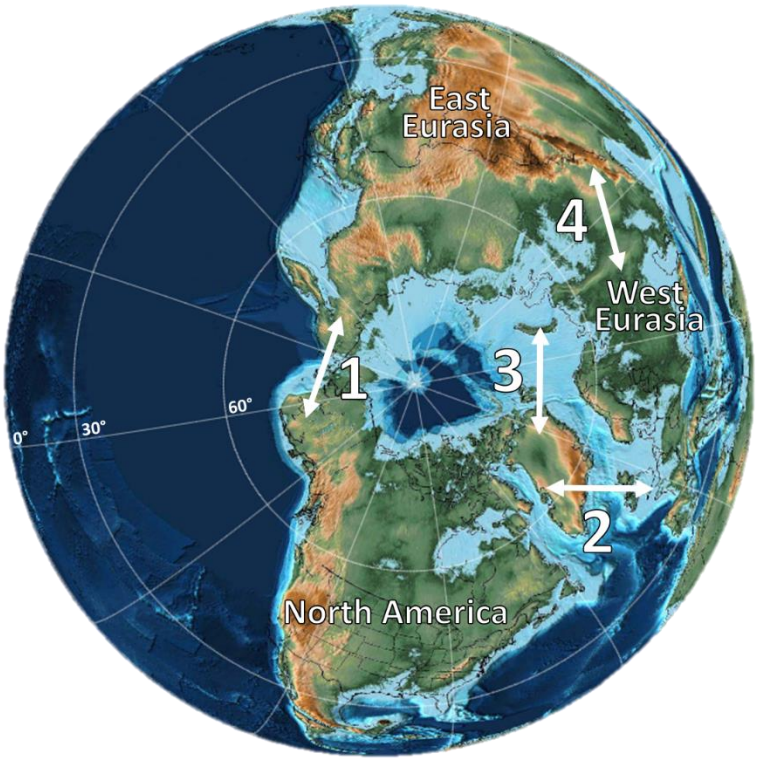


Figure 23: Possible dispersal routes across the Northern Hemisphere during the Paleocene: 1) Bering Land Bridge, 2) Thulean Route, 3) De Geer route and 4) Turgai Strait. Note that these routes were crossable at different times during the Paleocene. The palaeomap is sourced from Scotese (2016) and the figure is based on one presented in Brikiatis (2014).

The second peak in glires diversity occurred during the Oligocene Glaciations, which coincided with the diversification of Dipodoidea as well as other Asian myodont taxa, Afro-Asian glirids and North American sciurids and geomyoids. (Figure 19). The Antarctic ice sheet was expanding during this time (Zachos et al., 2001) and cooler global temperatures would have favoured rodents that could adapt, or that were preadapted to this weather. When rodent diversity contracted during the intense

warming of the Mid-Eocene Thermal Maximum, it is conceivable that cold-adapted populations would have retreated to cooler regions. The Tibetan Plateau, known as the 'Third Pole' (Qiu, 2008), would be an obvious choice of refuge for taxa in Asia and may have acted as source for species that emerged and thrived during the Oligocene Glaciations. This theory has relevance for the future of present day rodents and the global warming scenario we are currently experiencing. However, making inferences so far in the past is hard to prove, particularly with sampling biases of the fossil record. Despite this, there is plenty of evidence to show that the Tibetan Plateau acted as a refuge during recent glaciations, for which we have more precise data. Deng et al. (2011) were the first to propose the 'out of Tibet' hypothesis, in which Ice Age fauna did not emerge during the Pleistocene, but instead originated in the Pliocene of Tibet, from which they dispersed when the global temperatures dropped. This is now a widely recognised theory and has convincing evidence to support it from studies of the woolly rhinoceros (Deng et al., 2015), ancestral hyenas (Tseng et al., 2013a), the oldest big cat (Tseng et al., 2013b) and primitive blue sheep (Wang et al., 2015). It therefore seems likely that the Asian rodent groups to diversify during the Oligocene Glaciations, such as Myodonta, were already adapted to colder weather from residing in Central Asia. In fact, there is now evidence to suggest that dipodoids did indeed evolve in the Himalaya-Tibet and Central Asia regions during the Eocene and diversified out of this area when temperatures cooled off (Pisano et al., 2015). This result is supported by my study, which records the split between Dipodoidea and Muroidea during the Eocene and subsequent diversification of modern day dipodids in the Oligocene (Figure 19).

The next peak in Glires diversity occurs during the Miocene glacial expansion, following a sharp decline in diversity during the Late Oligocene Warming (Figure 19). This period coincided with major uplift in the Tibetan Plateau region, including the Tien Shan, Karakorum, Himalaya and Altai mountain ranges (Torsvik and Cocks, 2016). The rodent groups which diversified during this period include the Asian muroids and cricetids. Once again, it is easy to speculate that the Tibetan Plateau may have acted as a reservoir of cold-adapted species, which then spread out from the area when conditions across the continent became more favourable for them. However, there has also been research to suggest these mountain ranges at other times could have acted as a barrier to dispersal and that rodent movement was inwards, not outwards. Increasing amounts of Pliocene fauna have been sampled from these regions, which has resulted in better resolution of the fossil record. Li et al. (2015) used this detailed record of Pliocene fossils to show that cricetid rodents, rather than spreading out from Tibet during the Ice Age, travelled further inwards to colonise the Himalayan mountain range. Apparently they were not able to pass south of the mountains as no cricetid fossils could not be found on the other side. The hypsodont teeth of cricetids likely gave them an advantage when arid and open grasslands expanded, yet some taxa must have been pressured to retreat inwards from the plateau itself or

perhaps took the opportunity to exploit niches in even higher altitude regions. Further evidence for small mammals moving towards the mountain ranges encircling the plateau was supplied by Li and Wang's (2015) study on Pliocene zokors (spalacid mole-rats), and Polly et al's (2015) study on Pleistocene marmots. As Li and Wang (2015) pointed out, sampling deficiencies can bias our understanding of how the Tibetan Plateau shaped biota during glacial cycles. The current fossil record may suggest the movement of many large mammal groups 'out of Tibet' during cold snaps, but it remains to be seen if further sampling on the Himalayan side of the plateau truly supports this direction of travel. Either way, it would be difficult to prove that the plateau acted as a refuge for taxa during warming events or that it was a training ground for rodents entering the Oligocene or Miocene glaciations without this level of resolution, which becomes less obtainable the further back in time we go. With that being said, my data does not rule out the possibility that tectonic activity caused by uplift of the Tibetan Plateau was a potential driver of Glires evolution.

Conclusions

In this chapter I sought to disentangle the origins of Glires and to describe their biogeographic history. I have several results to support the importance of the Asian continent throughout the evolution of Glires. Firstly, my data provide evidence that Glires originated in Asia before spreading across the world. The order in which Sciuromorpha, Myomorpha and Ctenohystrica diversified remains unresolved. This is partly due to conflicting results from different types of data and the rapidity with which rodents first diversified in the Upper Cretaceous. As well as this potential trichotomy at the rodent root, the different types of data used for the total-evidence analyses altered the phylogenetic affinities of key rodent fossils. The placement of the Asian fossil *Tataromys,* for example, implied that Myomorpha + Ctenohystrica is derived from a ctenondactyloid-like ancestor in one analysis, and in a different analysis, that the whole of Rodentia is rooted on a ctenodactyloid-like ancestor. Unsurprisingly, these topological disparities filtered down into the analyses of biogeography. One analysis suggested a North American origin of Sciuromorpha and Myomorpha while the other pointed towards almost anywhere other than North America. This highlights the need to explore several types of data in total-evidence phylogenetic analyses and to adapt phylogenetic methods when large genomic data, such as UCEs, are used. Finally, my results point towards a potentially important role of the Tibetan Plateau in the evolution of Glires in Asia. The initial burst in Glires diversification coincides with the collision of the Indian plate into Asia and a further five peaks in Glires diversity occur during subsequent periods of uplift. Furthermore, some dips in rodent diversity are proceeded by recovery during glaciation periods and the diversification of some Central Asian taxa such as Myodonta. This

would seem to suggest that the Tibetan Plateau acted as a refuge and an evolutionary training ground for some taxa in the past, as we know it to have done so during the recent glaciations. A future role for the Tibetan Plateau as a refuge for rodents going through our current global warming scenario is speculative, but as the world's third largest 'pole', it could conceivably harbour less heat-tolerant species once more.

Chapter Five: anatomical sciuromorphy in 'protrogomorph' rodents

Abstract

Historically, high-level rodent taxonomy has been intertwined with the anatomical condition of the masseter muscle, leading to the widespread use of grades based on four anatomical conditions: protrogomorphy, sciuromorphy, hystricomorphy, and myomorphy. Although many previous studies have since shown these grades to be paraphyletic, the idea of a 'protrogomorph' rodent has remained popular for extinct species. We examined the oldest and most complete articulated skeleton yet known of *Ischyromys* (USNM 617532) from the late Duchesnean of West Canyon Creek, Wyoming. *Ischyromys* species are usually treated as anatomically protrogomorphous, but USNM 617532 shows attachment of the deep masseter anterior and dorsal to the infraorbital foramen on the rostrum, creating a zygomatic plate, and therefore exhibits anatomical sciuromorphy. A geometric morphometric analysis of cranial landmarks suggests that *Ischyromys typus* resembles extant, anatomically protrogomorphous rodents, whereas USNM 617532 falls within the range of non-protrogomorphous rodents. USNM 617532 differs from *Ischyromys typus* in terms of incisor procumbency, infraorbital foramen size and zygomatic arch shape. In addition to its distinctive rostrum and zygomatic plate, USNM 617532 possesses a sagittal crest, accessory cusps on the lower molars and large metaconules on the upper molars, features that help to diagnose this specimen as a member of *Ischyromys douglassi*. Masseteric patterns mapped onto a recent phylogenetic estimate suggest that none of the three main rodent clades, or even crown Rodentia itself, was characterized by an anatomically protrogomorphous common ancestor, although protrogomorphy did characterize many simplicidentate taxa on the stem leading to Rodentia.

Introduction

Several adaptations enable gnawing in rodents. They possess a single pair of ever-growing incisors in the upper and lower jaws, a reduced number of cheek teeth, a long diastema separating the incisors from the cheek teeth, and differentiation of the masseter muscle into three separate components to control jaw movement: the superficial masseter, deep masseter, and zygomaticomandibularis (terminology following Cox and Jeffery, 2011). The variation seen in the arrangement of the masseter tissues varies across rodents and falls into four morphologies: sciuromorphy, hystricomorphy, myomorphy, and protrogomorphy. These differences were first discussed by Waterhouse (1839) and then formally described by Brandt (1855). Protrogomorphy describes a condition where the origin of the masseter is limited to the zygomatic arch, a condition typically observed in non-rodents. Anatomical protrogomorphy is evident in many fossil rodents, but among living rodents only occurs in

Aplodontia (Figure 24a) and some bathyergoids (Landry 1957; Wood 1965; Maier and Schrenk 1987; Cox and Faulkes, 2014). The other three arrangements are characterised by divisions of the masseter extending anteriorly to varying degrees. In sciuromorphy (Figure 24b–c), the origin of the deep masseter extends anteriorly from the zygomatic arch, past the infraorbital foramen, onto the rostrum. The rostrum features a correspondingly broadened and angled margin of the zygoma and lateral rostrum, called the zygomatic plate, which in some taxa (e.g., geomyids, sciurids; Figure 24c), approaches the dorsal and anterior margins of the skull. In hystricomorphy (Figure 24d), the deep masseter remains restricted to the ventrum of the zygomatic arch and the zygomaticomandibularis extends onto the rostrum, passing through an enlarged infraorbital foramen along its course. Myomorphy (Figure 24e) is a combination of both sciuromorphy and hystricomorphy; both the deep masseter and zygomaticomandibularis spread anteriorly, the latter through the infraorbital foramen and the former onto the zygomatic plate.

Brandt (1855) used these morphologies to split the rodents into three groups: Sciuromorpha, Hystricomorpha, and Myomorpha. Zittel (1893) proposed another group called 'Protrogomorpha', into which he placed *Aplodontia* and many fossil forms. Rodent specialists have long recognised that these groupings, as described by anatomical terms, do not represent monophyletic clades (e.g., Wood, 1965; Hartenberger, 1985; Fabre et al., 2012). Unlike the morphological patterns of the upper jaw, the lateral (or anatomically hystricognathous; Tullberg, 1899; Wood, 1965) position of the angular process of the mandible in relation to the plane of the lower incisors has been proven to define the clade of caviomorph and hystricid rodents (i.e., Hystricognathi). This is opposed to a sciurognathous jaw, which indicates the position of the angular process in line with the lower incisor alveolus, as seen in all other rodents. Landry (1999) noted similarities of Hystricognathi with ctenodactylids and named this group Entodacrya, a term that has temporal priority over Ctenohystrica (Huchon et al., 2000), but is less widely known. Recent phylogenies (e.g., Marivaux et al., 2004; Churakov et al., 2010; Fabre et al., 2012; Asher et al., 2019; Swanson et al., 2019) support Ctenohystrica (i.e., guinea pig-related), squirrel-related, and mouse-related clades, with the masseter morphologies spread out across all three of these monophyletic groups.

The taxonomy of Brandt (1855) has priority over subsequent rodent classification schemes, and, arguably, should be at least partly retained to reflect broad consensus on the monophyly of each of these high-level clades. We, therefore, use Sciuromorpha and Myomorpha to indicate the squirrel- and mouse-related clades, and Ctenohystrica (Huchon, 2000) for the relatively novel clade including Hystricomorpha plus Ctenodactylidae. We also acknowledge the less well-known but prior suggestion of Entodacrya (Landry, 1999) for the clade of ctenodactylids-hystricomorphs; this term could serve as a total-clade cognate, encompassing crown Ctenohystrica and their stem relatives. When referring to

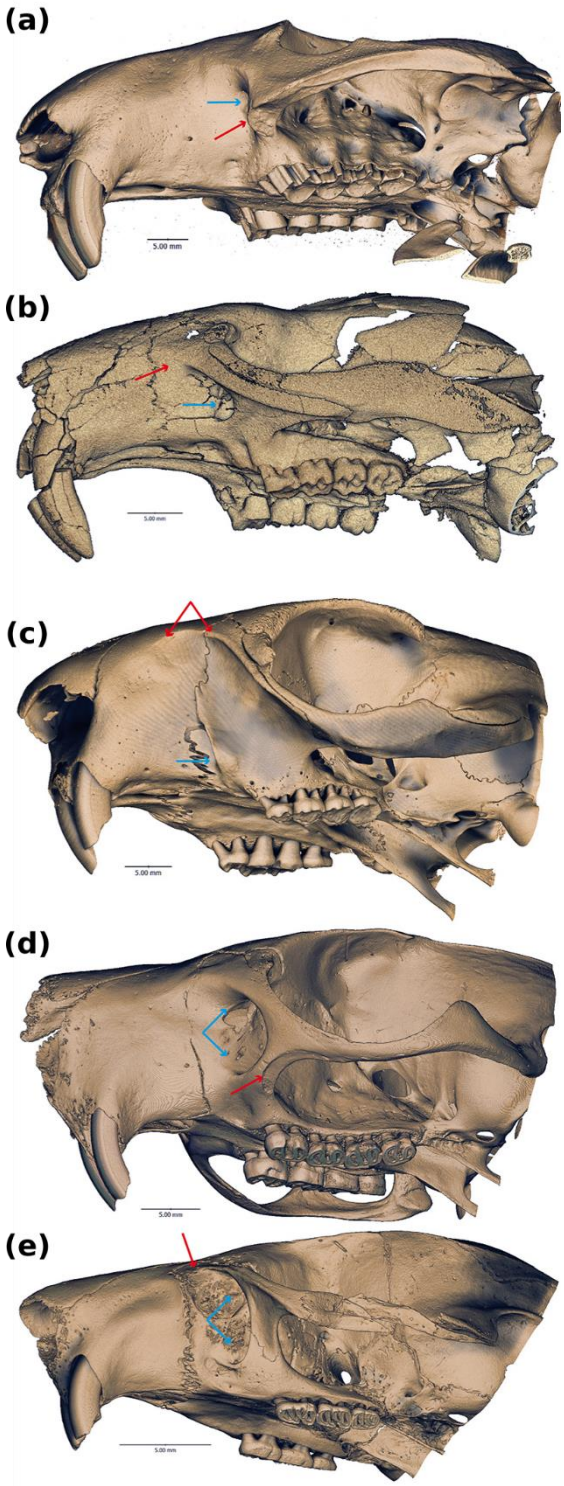


Figure 24: Anterior skulls of rodents in lateral views illustrating anatomical protrogomorphy (A, *Aplodontia rufa* UMZC E1861), sciuromorphy (B Ischyromys douglassi USNM 617532; C *Ratufa bicolor* UMZC E1570), hystricomorphy (D *Anomalurus beecrofti* UMZC E1414), and myomorphy (E *Platacanthomys lasiurus* UMZC E1917). Red arrows indicate anterior margin of deep masseter attachment; blue arrows indicate the infraorbital foramen. Scale bars equal 5 mm.

muscular types, we use the nouns sciuromorphy, myomorphy, hystricomorphy, and protrogomorphy, and the adjectives sciuromorphous, myomorphous, hystricomorphous, and protrogomorphous. Our use of these terms with suffixes '-morphy' and '-morphous' is purely anatomical and does not imply membership in any particular clade.

Authors such as Wood recognised homoplasy in masticatory types and argued that they represented 'adaptive levels' within Rodentia, involving changes in both cranial myology and dentition. His 'Grade One - protrogomorph radiation' included *Aplodontia* along with many extinct taxa such as Paleocene-Eocene Paramys, Eocene-Oligocene *Ischyromys*, and groups such as Cylindrodontidae and Mylagualidae. Furthermore, Wood disassociated Phiomyidae from these animals, believing it to be closer to thryonomyids. While he wrote that his “Protrogomorpha may be considered to be a clade” (1965, p. 118), he believed that all other rodent groups existed as grades. For example, his 'Grade Two' radiation was based on muscle and skull changes and contained anatomically sciuromorphous, hystricomorphous, and protrogomorphous species. His 'Grade Three' focused on hypsodonty and included species from the first two grades, which had developed ever-growing cheek teeth. This included species from groups such as ground squirrels and geomyoids. Wood considered the adaptations seen in 'Grade Two' to have arisen multiple times independently and, therefore, believed that the three classical suborders should make way for a correspondingly larger number of clades to represent such independent origins (Wood, 1965, p. 128).

'Ischyromyoidea', as defined by Wood (1965), comprises a paraphyletic assemblage of North American taxa. Historically, temporally diverse groups such as Paleocene and early Eocene *Paramys*, *Franimys*, *Reithroparamys*, and much younger, late Eocene-Oligocene *Ischyromys* have been associated with 'Ischyromyoidea' or 'Ischyromyidae' (Wood, 1965; Korth, 1994: figure 5.5). Following Korth (1994, p. 37), these genera have been regarded as anatomically protrogomorph: "the attachment for the masseter on the skull is limited to the ventral base of the zygoma (protrogomorphous) and the infraorbital foramen is small, opens anteriorly, and is not enlarged by muscle invasion or compressed by any modification of a zygomatic plate".

Wood spent many years reviewing these rodents (e.g., Wood, 1937, 1962, 1965, and 1976) and was the advisor of another eminent vertebrate palaeontologist, Craig Call Black. Both were experts in rodent anatomy and examined the same specimens. However, they did not agree on the supposed protrogomorphous nature of many fossil rodents. Black (1968) believed that all members of 'Protrogomorpha', including *Ischyromys*, were anatomically protrogomorphous. This meant that the term 'protrogomorph' was sometimes used taxonomically to denote members of what Black or Wood would have recognised as a natural group, as well as to describe the anatomical condition in which the masseter does not extend anteriorly or dorsally onto the rostrum from its attachment sites on the zygomatic arch.

In contrast to Black, Wood (1976) argued that some 'protrogomorph' ischyromyids were, in fact, anatomically sciuromorphous. Wood observed that some Chadronian specimens of *Ischyromys* had a more derived masticatory musculature than those from the Orellan (Wood, 1937, 1980) and referred these derived forms to a different genus, *Titanotheriomys*. Wood and Black's debate revolved around these two genera. Black (1968) synonymized *Titanotheriomys* with *Ischyromys*; Wood (1937, 1976) regarded them as closely related, but distinct. This was based heavily on Wood's myological interpretation of the crania. As summarised below, he (e.g., Wood, 1976) argued that at least some ischyromyines were anatomically sciuromorph, whereas Black (1968) regarded them all as protrogomorph.

Their disagreement, both anatomically and taxonomically, has yet to be resolved. Korth's (1994, p. 50) comprehensive treatment of North American fossil rodents concluded "it is most economical to include all ischyromyines in a single genus *Ischyromys*" pending a systematic revision of relevant material including juveniles. Heaton (1996: figure 14) undertook a cluster analysis of lower jaws, based on "mean population values for 83 characters: 25 on each molar and eight on the jaw" (1996, p. 391). His analysis yielded two clusters, corresponding to geologically older (largely Chadronian) and younger (Orellan and Whitneyan) groups of *Ischyromys*. He placed the Duchesnean West Canyon Creek

population among the taxa in the older cluster, along with one other Duchesnean population (Porvenir, Texas) and nine from the Chadronian, including *I. douglassi* from McCarty's Mountain (Montana). He thus concluded that the Duchesnean-Chadronian specimens of *Ischyromys* were morphometrically distinct from the younger populations, and used the subgenus *Ischryomys* (*Titanotheriomys*) for the former. We do not dispute this usage, but until a phylogenetic analysis is undertaken of all relevant species of ischyromyines (including many specimens unavailable to us), we follow Korth (1994) in using the generic name *Ischyromys* for these late Eocene and early Oligocene populations.

Wood (1976) focused on specimens from Montana, Wyoming, and Texas. From Montana, he examined material from Pipestone Springs and McCarty's Mountain (from the middle and early Chadronian, respectively; see Tabrum et al., 1996, Prothero and Emry, 2004). In Wyoming, the major sites of interest were Beaver Divide (White River Formation, Chadronian; Emry, 1975) and Bates Hole (early to middle Chadronian). Most Texan specimens he studied came from Ash Spring (middle Chadronian). He observed that some specimens of *Ischyromys* had a tilted zygomatic plate that sloped from the maxillary root of the zygoma to the infraorbital foramen (Wood, 1976). Furthermore, he observed a crest that curved medially across the plate, separating the fossa of the deep masseter from the incline below the infraorbital foramen (Wood, 1976: his figure 2). To Wood, this suggested that there was some migration of what he called the "masseter lateralis" (e.g., Wood, 1976: figures 3b, 3c, referred to as the deep masseter here following Cox and Jeffery, 2011) off the ventral surface of the zygoma, signalling important changes had occurred in the gnawing mechanism. Black agreed to some anterior migration of the deep masseter and slight compression of the infraorbital foramen in some specimens of *Ischyromys*. However, he still considered specimens in this genus to be anatomically protrogomorph (Black, 1968).

Wood described many *Ischyromys* specimens (his *Titanotheriomys*) as anatomically sciuromorph, without crests cutting across the zygomatic plate denoting the anterior end of the deep masseter, but with the dorsal limit of this muscle extended forwards, anterodorsal to the infraorbital foramen (Wood, 1976: figure 3). In contrast, Black (1968, pp. 279–282) argued that specimens from Beaver Divide were distorted and that Wood's observations resulted from this distortion and a failure to recognise morphological variation within the population. Wood, however, was not convinced by the latter and claimed that no Chadronian collection from a restricted spatio-temporal location showed a transitional series of forms, as expected if there were variation within a population. As recounted by Wood (1976), both men were civil, meeting in person to look at the specimens together and yet, ultimately, their debate was never resolved. Wood's idea of anatomical sciuromorphy within 'Ischyromyidae' did not play a major role in subsequent literature, and *Ischyromys* (including

Titanotheriomys) has since been described as 'protrogomorph', at least in a taxonomic sense (e.g., Dawson, 1977; Vianey-Liaud, 1985; Korth, 1994).

Figure 25: CT scan of USNM 617532 embedded in matrix (right) and virtually dissected (left). Scale bar equal 50 mm.

In this study, we describe one of the oldest and best-preserved skeletons of *Ischyromys* (USNM 617532, Figure 25) and consider its evolution in the context of a recent phylogenetic study (Asher et al., 2019). This fossil was collected by Robert Emry in 1971 and is from the White River Fm of West Canyon Creek, Wyoming. It is Duchesnean in age (i.e., 40–37 Ma, Prothero and Emry, 1996) and among the geologically oldest localities from which ischyromyines (including Wood's *Titanotheriomys*) are known. The Duchesnean River Formation of Utah contains relatively few taxa and is distinct from the Uintan fauna, but more similar to the Chadronian fauna (Emry, 1981). The Duchenean age of West Canyon Creek is well supported by the co-occurrence of two rare artiodactyls (*Agriochoerus maximus* and *Brachyhyops*), which are found together in only three locations: West Canyon Creek itself, lower sections of Flagstaff Rim (Wyoming) and in the Lawpoint Member of the Duchesnean River Formation in Utah (Emry, 1981). We consider the USNM 617532 skeleton to belong to *Ischyromys douglassi*, a taxon originally described from McCarty's Mountain, Montana (Black, 1968). Based on the presence

of *Ardynomys occidentalis* in both the Porvenir fauna of west Texas and the McCarty's Mountain fauna (and nowhere else), and on the similarity of *I. blacki* from the Porvenir and *I. douglassi* from McCarty's Mountain, Wood (1974) concluded that the two faunas were close in age, with the latter perhaps being slightly younger. The Porvenir fauna is widely regarded as late Duchesnean in age (Prothero, 1996). Several other taxa of the West Canyon Creek fauna suggest that it is a close correlate of the Porvenir and McCarty's Mountain faunas, perhaps closer to the former than the latter. Thus, we regard the specimens of *I. douglassi* discussed herein as being of late Duchesnean age. Heaton's (1996) analysis of dental evolution in *Ischyromys* regarded this West Canyon Creek species as new but did not name it. His analysis focused on the lower dentitions of ischyromyines, including jaws and teeth from West Canyon Creek. The dental crown morphology of USNM 617532 remains occluded by matrix, but is nonetheless visible with computerized tomography (CT) and thus amenable to comparison to other West Canyon Creek specimens. We also compare USNM 617532 and other West Canyon Creek specimens to younger, Orellan (34-32 Ma) species of *Ischyromys*, including *I. typus* (USNM 16828, AMNH 144628, ROM V1007 and BMNH PV M 7855) and note previously missed details of the *Ischyromys* ear region. Finally, we provide a geometric morphometric assessment of the masticatory configuration of these specimens in light of Wood and Black's debate.

Methods

Institutional acronyms

AMNH, American Museum of Natural History (New York); BMNH, The Natural History Museum (London); ROM, Royal Ontario Museum (Toronto); USNM, National Museum of Natural History (Washington, D.C.).

Taxon sample

We obtained microCT scans of 39 rodent and two lagomorph skulls (Appendix 4a), chosen to include representatives of each of the four anatomical conditions of the masseter muscle from the three major extant rodent groups and a number of specimens representing four fossil taxa. These were *Gomphos elkema* (MAE BU14467, a mimotonid from the early Eocene Bumbanian Asian Land Mammal Age), *Paradelomys crusafonti* (UM-ACQ-6618, a theridomyid from the middle Eocene following Vianey-Liaud and Marivaux, 2016), *Ischyromys* from the Orellan Brule Formation (USNM 16828 from Wyoming, ROM V1007 from Nebraska, AMNH 144628 from Stark County North Dakota, CMNH 588 from Badland Creek Nebraska, and BMNH PV M 7855 from the White River Fm of South Dakota), and USNM 617532 from the Duchesnean of West Canyon Creek, Wyoming. Computerized tomography

scans of two of our USNM skulls (617532 and 16828) are available via morphosource.org project 945 (https://www.morphosource.org/Detail/ProjectDetail/Show/project_id/945). We considered the anatomy of several additional *Ischyromys* specimens, also from West Canyon Creek, from a collection housed at the USNM. Nomenclature used for dental morphology is depicted in Figure 3 and follows Wood and Wilson (1936) and Marivaux et al. (2004).

Geometric morphometrics and character state reconstructions

We made virtual reconstructions with Drishti v.2.6.4 (Limaye, 2012; Figure 27a) and identified 21 digital landmarks in 3D based on Morris et al. (2018) from the right side of each cranium (Appendix 4b; Figure 27b). Of the five *Ischyromys typus* specimens, only USNM 16828 preserved the structures required for the landmarking. The effects of specimen size, position, and orientation were eliminated by performing a Procrustes fit on the superimposed sets of landmarks using MorphoJ (Klingenberg, 2011). The locations of the specimens across morphospace were then visualised by performing a principal component analysis (PCA). The anatomical condition of the masseter for each specimen of an extant species, based on Maier and Schrenk (1987), Hautier et al. (2008), and Cox and Faulkes (2014), was then superimposed onto the PCA results to be able to observe overlap among fossils and non-rodents in our sample. In addition, we used the phylogeny of Asher et al. (2019: figure 27) to reconstruct the evolution of rodent masticatory types using parsimony ancestral state reconstruction in PAUP 4.0a build 167 (Swofford, 2003) and Mesquite v3.51 (Maddison and Maddison, 2019), also taking into account character optimizations in fossil taxa as described by Wang (1997), Meng et al. (2003), Marivaux et al. (2004), Wible et al. (2005), and this chapter.

Results

Description

Ischyromys douglassi USNM 617532 (Figure 25) is embedded in matrix and consists of a complete skull with a dorsally fragmented braincase and upper and lower dentitions, plus a nearly complete skeleton. The dentition consists of the right and left dP4, M1–3 and dp4, m1–3, with alveoli for upper right and left deciduous and permanent P3 loci. The skeleton consists of cervical, anterior thoracic, and posterior lumbar vertebrae, partially articulated but displaced ribs, sacrum, fragmentary os coxae with acetabulum present only on left side, partial left scapula with articulated left humerus, radius, ulna and partial manus, glenoid fossa of right scapula and damaged right humerus and proximal ulna, articulated left femur, tibia, fibula, and broken but articulated left astragalus and calcaneum (missing their posterior aspect), distal right femur and proximal tibia, and disarticulated phalanges and some

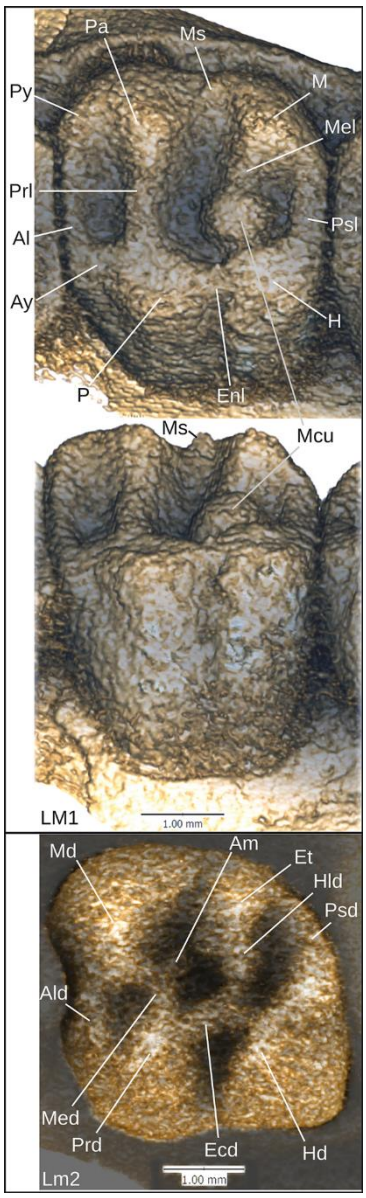


Figure 26: Occlusal (top), occlusal-lingual (middle) morphology of upper M1 and occlusal morphology of lower m2 (bottom) of *Ischyromys douglassi* (USNM 617532). Uppers: Ay, anterostyle; Enl, endoloph; H, hypocone; M, metacone; Mcu, metaconule; Mel, metaloph; Ms, mesostyle; P, protocone; Pa, paracone; Prl, protoloph; Psl, posteroloph; Py, parastyle. Lowers: Ald, anterolophulid; Am, anterior medial ('accessory') cusp; Ecd, ectolophid; Et, entoconid; Hd, hypoconid; Hld, hypolophulid; Med, metalophulid; Md, metaconid; Prd, protoconid; Psd, posterolophid; Psl, posteroloph. Scale bars equal 1 mm.

caudal vertebrae. Several additional specimens from West Canyon Creek further illustrate the anatomy of *I. douglassi*, including skulls with at least intact rostra and some associated upper and lower dentitions (USNM 475454 and 489144 illustrated in Figure 28, plus USNM 475451, 475453, 475457, 475458, and 489148).

Upper dentition

USNM 617532 has one deciduous premolar (dP4) and three molars on each side of the upper jaw (Figure 29a–f). Although P3s and permanent P4s are missing in the maxilla of this specimen, dorsal and anterior to each dP4 are empty alveoli with sufficient space on each side to house both a P3 root and P4 crown. We therefore infer for this specimen the typical *Ischyromys* dental formula of 1.0.2.3/1.0.1.3. The dP4 is heavily worn and lacks detail of its occlusal anatomy, such as transverse crests or metaconules that may have been present. The dP4 differs from the molars in having no hypostria (vertical grooves) on the lingual surface of the tooth. The paracone and metacone are still visible; the protocone and hypocone are almost entirely worn down.

Molars show a large metaconule on the lingual end of the metaloph in M1–3, which dams the posterolingual valley (Figure 3, Figure 29). M1 and M2 are quadritubercular with pronounced hypocones. There is a distinct metacone, metaconule, and hypocone, with very weak development of the metaloph connecting them. The anterior arm of the paracone is extended to form a small parastyle. The anterior arm of the protocone is also extended, forming a small anterostyle, which is connected to the parastyle by an anteroloph. The posterior end of the protocone and anterior end of the hypocone are similarly extended, forming the endoloph that joins them. The posterior arm of the hypocone extends lingually, producing a prominent posteroloph. A very small mesostyle is present on M1 between the paracone and metacone, joined by the posterior and anterior arms of these cusps respectively. The M3 is similar and also shows a prominent mesostyle, but differs from the other molars in having an obliquely

oriented metaloph and consequently short transverse dimension of the posterior side of the tooth. The molars feature a metaconule on the lingual end of the metaloph, well developed on M1–3, which almost dams the entire posterior lingual valley. Due to the oblique metaloph on M3, the metaconule shows a very slight metacone-protocone connection. Large upper molar metaconules are reported in *Ischyromys blacki, I. douglassi* (Figure 26, Figure 28C), and *I. junctus* (Wood, 1980) but not species of *Ischyromys* such as *I. typus* (Figure 29i).

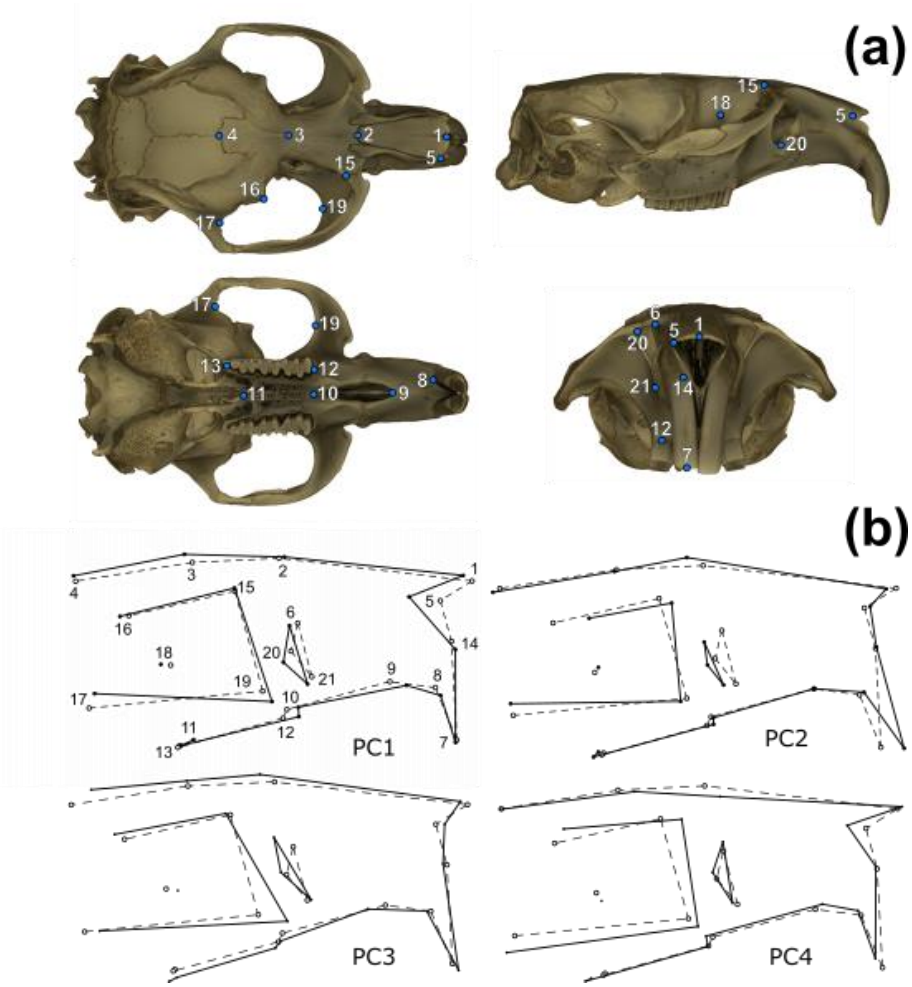


Figure 27: a) Graphic depictions of 3D landmarks on a skull of *Arvicola amphibius* (UMZC E2805), b) Wireframe graphs summarising the cranial shape variation explained by principal components (PCs) 1–4. The dotted line wireframe represents the mean cranial shape across the samples and the solid line wireframe represents the magnitude of shape change for a particular principal component (scale factor 0.1). The landmarks are numbered with descriptions provided in Appendix 4b.

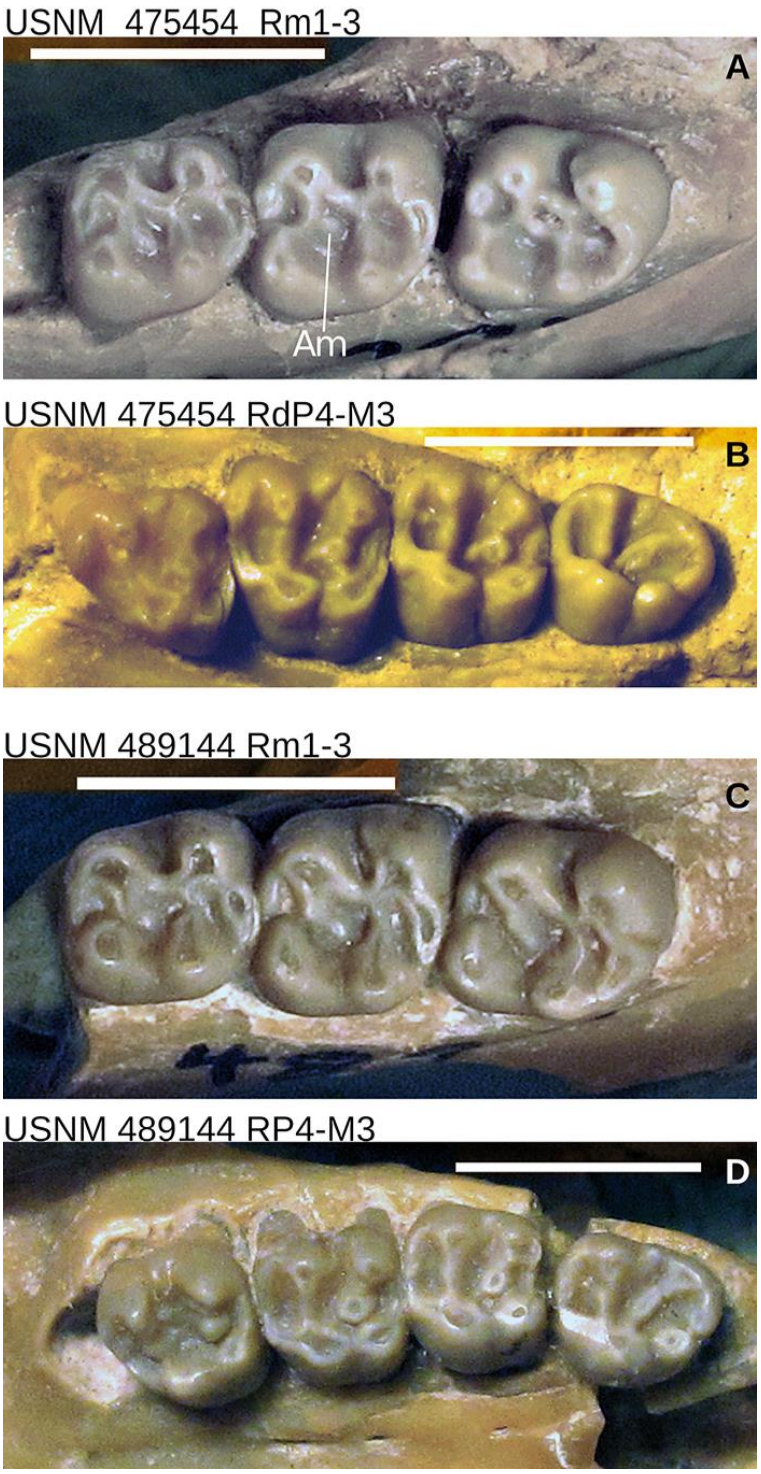


Figure 28: Lower (A, C) and upper (B, D) cheek teeth of *Ischyromys douglassi* from West Canyon Creek, Wyoming. Anterior is to left, buccal top. Am = anterior medial ('accessory') cusp. Scale bars equal 5 mm.

Lower dentition

The lower dp4 of USNM 617532 is also heavily worn (Figure 29e–f), with only a trace of the metaconid and hypoconid remaining. This tooth is buccolingually narrower, particularly anteriorly, than the square-shaped molars. The lower molars possess a trigonid consisting of the metaconid and protoconid. The two cusps are connected by the metalophid, which curves posteriorly from the metaconid to the protoconid. The posterior arm of the metaconid is well developed, almost reaching the entoconid. The anterior arm of the protoconid is extended, forming an “anterolophulid” according

to Marivaux et al. (2004: figure 1). The talonid is wider than the trigonid and comprises a large hypoconid, smaller entoconid, and smaller still, a hypoconulid. The hypoconid is connected to the protoconid via a curved ectolophid but no mesoconid is observable on this crest. The hypolophid intersects this crest, connecting the entoconid to the hypoconid. The posterior arm of the hypoconid extends medially to form a pronounced posterolophid that reaches the lingual side of the tooth. A distinct anterior medial cusp originates on the posterior edge of the metalophid (labelled 'Am' on Figure 26, Figure 28) and almost divides the lingual valley. This cusp is best observed on m2 and m3 where the cusp almost extends as far as the entoconid, splitting the anterior lingual valley in two. Heaton illustrated this cusp and referred to it in his text as the "anterior medial accessory cusp" (Heaton, 1996, his figure 1 and p. 375). The lower molars of USNM 617532 are larger than those of *Ischyromys blacki* and smaller than those of *I. typus*. Its lower m1 is slightly shorter, and its m2 and m3 lengths are within the range for *I. douglassi* given by Black (1968: table 2). The m3 of USNM 617532 is larger than that of *I. veterior*, but its m1–2 are slightly shorter in length and broader in width. Overall, its dental lengths are close to those recorded for *I. douglassi* as reported by Black (1968: table 2), although some are near or slightly below the lower end of his ranges.

All of the *Ischyromys typus* specimens we examined date to the Orellan (34-32 Ma) and have the typical *Ischyromys* dental formula of 1.0.2.3/1.0.1.3, although only USNM 16828 (Figure 30e–h) and BNMH PV M 7855 are sufficiently complete to retain a full complement of teeth. The P3 is a small peg-like tooth with a cone and partially developed peripheral cingulum (Figure 29h–i). The P4 is molariform with a pronounced hypocone, protocone, paracone, and metacone. The paracone and protocone are connected by a straight transverse crest, the protoloph. The metacone and hypocone are likewise joined by the metaloph (Figure 26i). The P4 is distinguished from the molars by the absence of the hypostria between the protocone and hypocone, on the lingual surface of the tooth. Neither paraconule nor metaconule are present. Anterolophs and posterolophs are present, although the former is weak and terminates just lingual to the parastyle. As for P4, the molars are quadritubercular and have pronounced hypocones. From M1 to M2 to M3, the metaloph becomes progressively shorter and more oblique, such that the M3 is posteriorly narrower than the M2, which is posteriorly narrower than M1. The anteroloph is more developed in the molars than in P4 and extends to the lingual margin of the tooth.

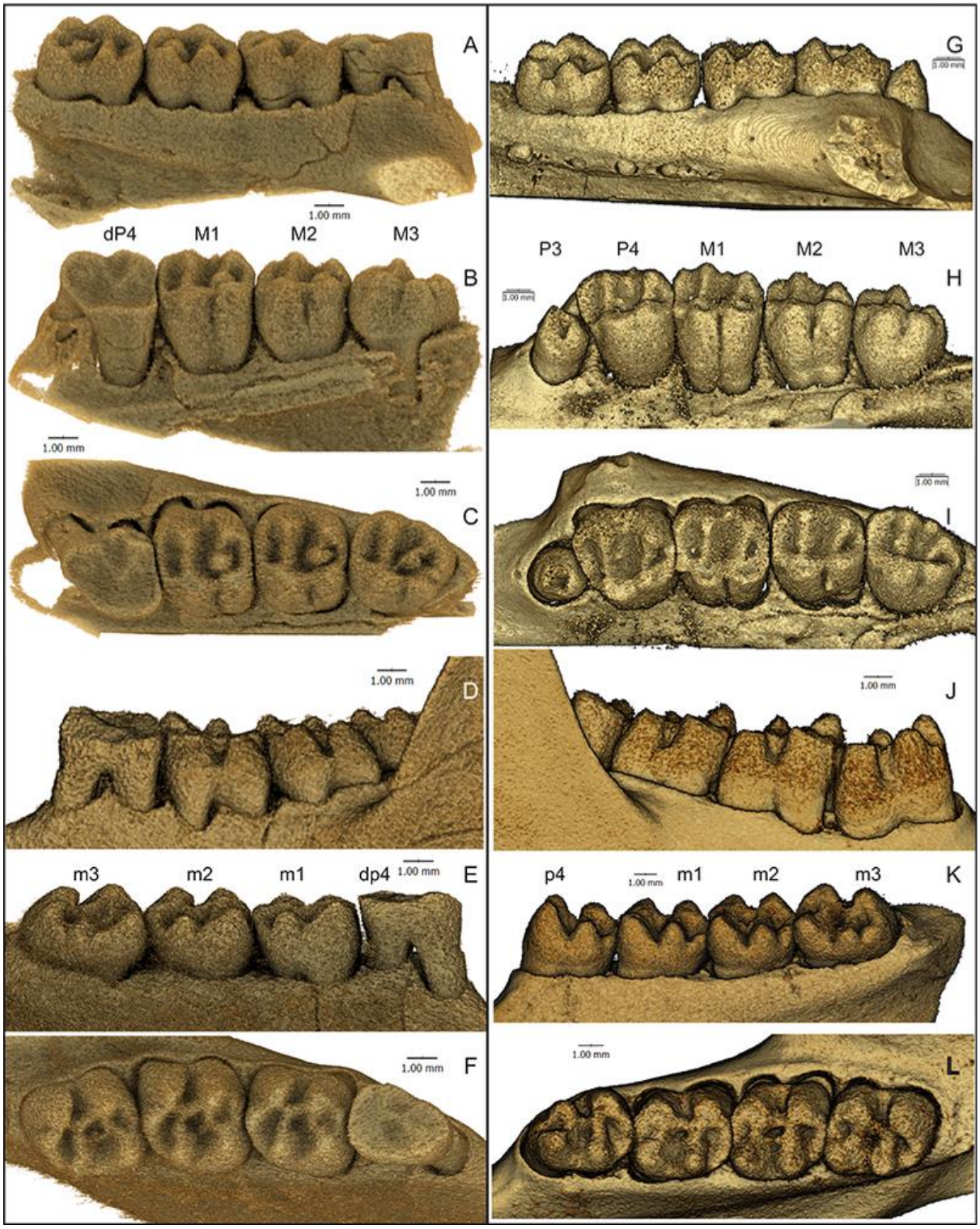


Figure 29: Left Upper (A–C, G–I) and left (D–F) and right lower (J–L) dentitions of *Ischyromys douglassi* USNM 617532 (A–F) and *I. typus* USNM 16828 (G–L). A, D, G, J show buccal, B, E, H, K show lingual, and C, F, I, L show occlusal views. Scale bars equal 1 mm.

The teeth of the lower jaw of *Ischyromys typus* (e.g., USNM 16828) are also lophate (Figure 29j–l). The p4 has a small trigonid, appearing reduced anteriorly in comparison to m1 and m2. The m3 is extended and rectangular in shape. The lower teeth are fairly simple with a narrow trigonid, featuring a metaconid and protoconid. These two cusps are connected by a metalophid, which curves posteriorly from the metaconid towards the protoconid. The anterior side of the protoconid extends slightly to form a short anterolophid. This is best seen in the slightly more worn p4 and m1. The talonid is wider

than the trigonid (especially in p4), which has a large hypoconid and smaller entoconid. A hypolophid connects these two cusps. The hypoconid is joined to the protoconid via a curved ectolophid, which lacks a mesoconid. A well-developed posterolophid runs from the hypoconid to the lingual side of the tooth. There is no distinct hypoconulid.

Skull

The skull of USNM 617532 is broad and the rostrum comes to a blunt end. This and other specimens (see below) show that the species is anatomically sciuromorphous, consistent with the interpretations of Wood (1976). USNM 617532 and other West Canyon Creek specimens exhibit a scar of the deep masseter extending anterodorsally from the ventral surface of the zygoma, passing above and anterior to the infraorbital foramen and terminating close to the maxillary-premaxillary suture, contributing to a zygomatic plate on the lateral surface of the rostrum (Figure 24b, Figure 30b, d).

USNM 457454 from West Canyon Creek also shows a parenthesis-shaped scar on the bottom of the arch, which curves inward toward the P4. This distinct scar is also present in *Ischyromys typus* (Figure 30h); however, unlike *I. typus*, it shows a rounded but prominent ridge that continues forward and upward beyond the infraorbital foramen to merge with the dorso-lateral surface of the rostrum, with an anterior extent at or near the maxilla-premaxilla suture. Other West Canyon Creek specimens with a well-preserved rostrum (USNM 457453, 475757, 475458) show a similar rostral morphology as in USNM 617532 (Figure 30b–d), except for one particularly young specimen (USNM 489148), which retains dP3–dP4 and lacks any sign of a mineralized M3. Here, the scar is relatively weak with a termination slightly posterior to the premaxilla-maxilla suture.

Although USNM 617532 has jaw closing musculature resembling specimens referred to *Titanotheriomys* by Wood (1976), it shows several differences. *Titanotheriomys* was described by Wood (1976) as having an arched dorsal profile of the skull; this is relatively flat in USNM 617532. Furthermore, Wood characterized *Titanotheriomys* as lacking confluent supraorbital crests. In most *Titanotheriomys* specimens Wood examined, the supraorbital crests did not converge but showed a lyrate shaped space dividing them. In USNM 617532, however, the two supraorbital crests unite at the frontal suture and join to form a single sagittal crest (Figure 30a), as in other West Canyon Creek specimens (e.g., USNM 475458).

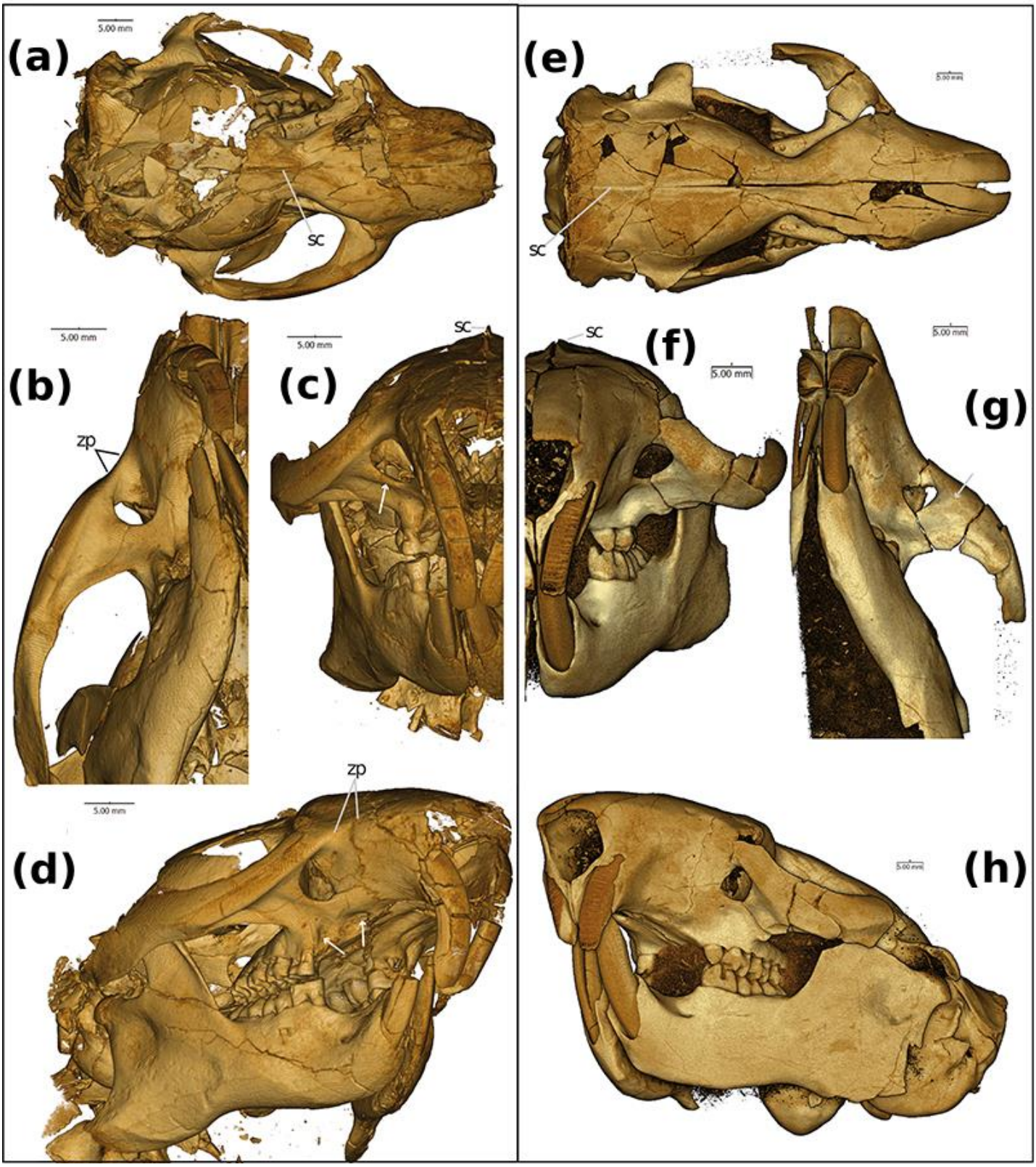


Figure 30: Skulls of *Ischyromys douglassi* USNM 617532 (A–D) and *I. typus* USNM 16828 (E–H). Arrows in C and G show attachments of deep masseter ventral to the infraorbital foramen (C); white arrows in D show fossae for the attachment of the buccinator muscle (left) and tendon of the superficial masseter (right). 'sc' = sagittal crest, 'zp' = zygomatic plate. Scale bars equal 5 mm.

The zygoma and rostrum of USNM 617532 differ from those of *Ischyromys typus* in a number of ways. From a dorsal view, the skull appears proportionally wider than that of *I. typus* (e.g., USNM 16828, Figure 30). The rostrum is also proportionally shorter than in *I. typus* and rather than tapering to a point (Figure 30e), it is blunt (Figure 30a). These two features of USNM 617532 bear a superficial resemblance to *Aplodontia rufa*. The maxillary root of the zygoma is broader in USNM 617532 than in *I. typus*, and there are two notable scars which, based on comparisons with extant rodents, indicate attachment sites of the deep masseter. As in *I. typus*, there is a scar which curves medially at the anterior end of the zygoma. However, unlike *I. typus*, this scar is very short and weak. The area between the dorsal edge of this scar and the infraorbital foramen is very inset, giving the ventral edge of the foramen a pronounced lip (Figure 30c). Furthermore, a second scar of the deep masseter can be observed to extend from the ventral surface of the zygoma. This scar follows the dorsal border of the maxillary root, ascending at a 45 degree angle. It reaches past the dorsal border of the infraorbital foramen, continuing upwards until it curves antero-medially to reach the maxillary-premaxillary suture. These features define a zygomatic plate (Figure 24b; 'zp' in Figure 30b, d), with its dorsal limit superior to the infraorbital foramen. The muscle scars indicate that the deep masseter originated on the rostrum, anterior to and above the infraorbital foramen, as well as covering the surface area below the infraorbital foramen. These scars demonstrate that *I. douglassi* is anatomically sciuromorphous, as described by Wood (1976) in *I. veterior* (referred to by Wood as '*Titanotheriomys wyomingensis*'). The infraorbital foramen, although much smaller than those seen in anatomically hystricomorphous rodents (Figure 24d), is proportionally larger than the foramina seen in protrogomorphous rodents, such as *Aplodontia rufa*. As in other specimens of *Ischyromys* (e.g., *I. typus*, Figure 30f) the foramen in *I. douglassi* narrows towards its ventromedial edge (Figure 30C). Other notable features on the rostrum of *I. douglassi* are two well-defined pits, anterior to each upper toothrow and ventral to the infraorbital foramen (Figure 30d). One of these pits is a small but deep depression antero-lateral to the alveolus of the P3. Following Wood (1976), this is likely to be a fossa for the attachment of the buccinator muscle. The other larger and more pronounced depression lies anteroventrally to the infraorbital foramen. This marks the attachment site of the tendon of the superficial masseter. The site of attachment for the tendon appears particularly large in USNM 617532 (Figure 30d) relative to USNM 16828 (Figure 30h).

The general features of the rostrum in *Ischryomys typus* agree with the descriptions provided by Wood (1937) and Black (1968) of various *I. typus* skulls found in the White River Formation. There is a clear muscle scar for the deep masseter on the ventral surface of the zygoma (Figure 30f–h). The scar curves medially at the anterior end of the zygoma, and continues until the maxillary root of the zygoma, marking the anterior limit of the muscle. As described by Wood (1976), the anterior surface of the

maxillary root of the zygoma in *I. typus* is tilted anteriorly, from the scar of the deep masseter to just below the infraorbital foramen (Figure 30h). Wood described this area as a "diagonal zygomatic plate" (Wood, 1976: p.252), although there is no evidence on the specimens examined in this study to suggest that the deep masseter attached here. Both *I. douglassi* (USNM 617532, Figure 30c) and *I. typus* (USNM 16828, Figure 30f) show large infraorbital foramina, mediolaterally wider than dorsoventrally tall. Wood (1976) described a depression for the tendon of the superficial masseter anterior and lateral to the P3 on *I. typus*. Although there are no rugose areas observable on USNM 16828 (Figure 30h), there is a distinct pit for the tendon of the superficial masseter in AMNH 144628. These specimens of *I. typus* are consistent with the generally held view that *I. typus* was anatomically protrogomorphous. Additionally, the angle between the coronoid process relative to the antero-posterior axis of the mandibular ramus is slightly smaller in USNM 617532 (about 114 degrees; Figure 30d) compared to *I. typus* (about 121 degrees in USNM 16828; see Figure 30h).

The inner ear region of USNM 617532 is well preserved. Wahlert (1974, p. 393) writes that based on absence of a stapedial foramen, the stapedial artery is absent in *Ischyromys*. Bertrand and Silcox (2016, p. 12) cite Wahlert and note further that "in ROM V1007 and AMNH 144628, the cast of the stapedial artery is absent". However, the carotid foramen, evident along the medial aspect of the auditory bulla, shows a connection with a substantial groove on the ventrum of the promontorium of the periotic (or 'promontory') in all CT scanned specimens of *Ischyromys* examined so far, including USNM 617532 (Figure 31a–c; see also Asher et al. 2019: figure 2), ROM V1007 (Figure 31d–f), USNM 16828 and AMNH 144628. This is further illustrated in media #M578238 of morphobank.org project 2769 (Asher et al., 2019). This groove traverses the fenestra vestibuli (Figure 31b, e), is continuous with the internal carotid foramen and canal proximally (Figure 31c, e, f), and leads to a gap in the roof of the epitympanic recess and adjacent foramina distally (Figure 31f), consistent with the presence of a patent stapedial artery. Given the size of the promontory groove (Figure 31b, c, e, f), which is comparable to those seen in *Exmus* and *Cocomys* (Li et al., 1989: figures 3, 4; Wible et al., 2005: figures 6, 7), we regard the presence of an artery as more likely than the presence of a nerve or vein only. Black (1968: figure 19) mistakenly labelled a "stapedial" foramen on the medial aspect of the auditory bulla in a specimen of *I. douglassi* (CMNH 1122), but the structure in question is actually a jugular (also known as posterior lacerate) foramen, anterolateral to the hypoglossal foramen and posterolateral to the carotid foramen. These three structures are correctly labelled for a specimen of *I. typus* (AMNH 694) in Wahlert (1974: figure 9).

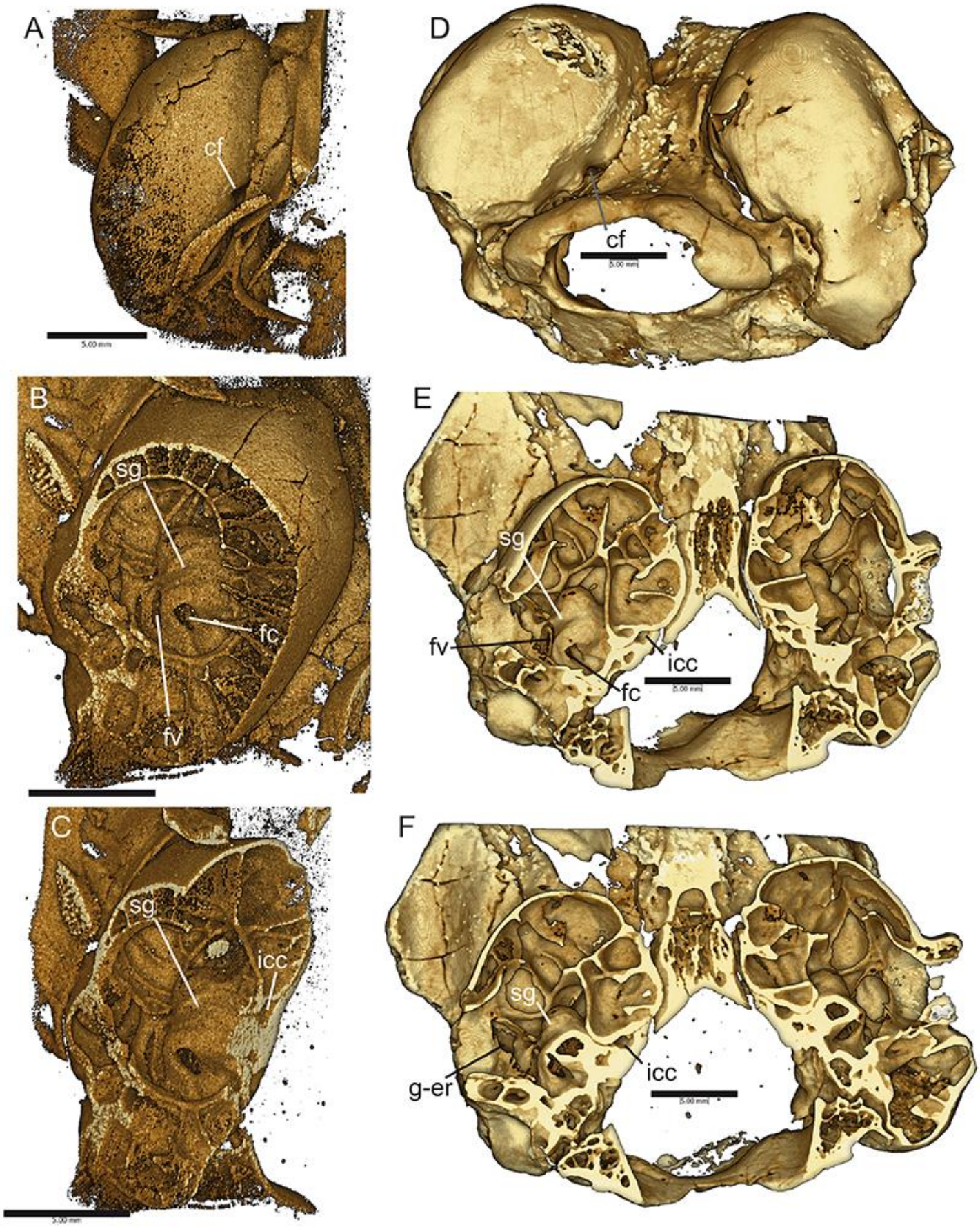


Figure 31: Ear region and associated foramina in *Ischyromys douglassi* (A–C, USNM 617532) and *I. typus* (D–F, ROM V1007). cf = carotid foramen, fc = fenestra cochleae, fv = fenestra vestibuli, icc = internal carotid canal, g-er = gap in epitympanic recess, sg = stapedial groove. Scale bars equal 5 mm.

Geometric morphometric analysis and character evolution

The PCA analysis of the 21 3D cranial landmarks (Figure 27) shows that 90% of the shape variation is explained by the first 14 principal components and just over 50% of variation is explained by the first four (see Appendix 4c for all eigenvalues). Interpretation of wireframe graphs indicate several cranial shapes changes associated with PC1 (17% of variation; Figure 32). Increasingly positive values along PC1 represent a shorter length of the naso-premaxillary suture in relation to the internasal suture, an increasingly anterior limit of the incisive foramina, and increased elevation of the posterior root of the zygoma in relation to the anterior root. PC1 also shows the infraorbital foramen increasing in diameter. The negative values along PC1 capture cranial shape changes associated with the presence of a large zygomatic plate and the positive values reflect extreme anatomical hystricomorphy. Increasingly positive values of PC2 (13.6% of variation) represent a trend towards increasing incisor procumbency, increasing length of the naso-premaxillary suture in relation to the internasal suture, and decreasing infraorbital foramen size. PC2 also describes the position of the anterior limit of the interior zygomatic arch edge. For positive values of PC2, the interior edge is in line with the orbital margin; for negative values, the anterior limit extends anteriorly past the orbital margin. PC3 (11.6% of variation) separates the extant rodents and fossils from the lagomorph *Lepus*. *Lepus* represents the most negative value along the PC3 axis, associated with a smaller infraorbital foramen, shorter naso-premaxillary suture, decreasing incisor procumbency, and an increasingly anterior limit of the frontal suture as it crosses the orbital margin. Negative values of PC4 (9.7% of variation) also indicate a smaller infraorbital foramen. The remaining 48.2% of cranial shape variation is explained by PC5–35, each accounting for progressively less variation.

When comparing the locations of species in morphospace across the first two principal components (30.5% of variation), anatomically sciuromorphous species form a wide cluster (Figure 32). Hystricomorphous species form an equally large cluster, which overlaps with the sciuromorphous and the more compact myomorphous cluster at the centre of morphospace. Anatomically protrogomorphous species form a small cluster that overlaps slightly with the myomorphous cluster and even less with the hystricomorphous cluster. *Ischyromys douglassi* (USNM 617532) is centrally placed and nested within the overlap between the hystricomorphous and sciuromorphous species. *Ischyromys typus* (USNM 16828) lies on the margin of the protrogomorphous cluster (Figure 32), closer to *Aplodontia* and bathyergoids than any other rodent based on the first two principal components. In contrast, and as noted above, *I.douglassi* is anatomically sciuromorph by virtue of its zygomatic plate defined in part by the masseter scar anterior and dorsal to the infraorbital foramen (Figure 24, Figure 30). In terms of shape space, it appears in a region in which two extant, non-protrogomorph clusters overlap (Figure 32), and is separated from *I. typus* by a larger distance on PC2 than PC1.

Ischyromys typus occupies a more positive position on the PC2 axis than *I. douglassi* indicating that in comparison, the latter has less profound incisor procumbency, a larger infraorbital foramen, a shorter naso-premaxillary suture, and that the interior edge of the zygomatic arch is anterior to the frontal suture as it crosses the orbital margin. *Paradelomys crusafonti* (Vianey-Liaud and Marivaux, 2016) is an anatomically hystricomorphous theridomyid and lies within the overlap of hystricomorphous and sciuromorphous clusters. *Gomphos*, a mimotonid stem-lagomorph, also lies within this region of overlap.

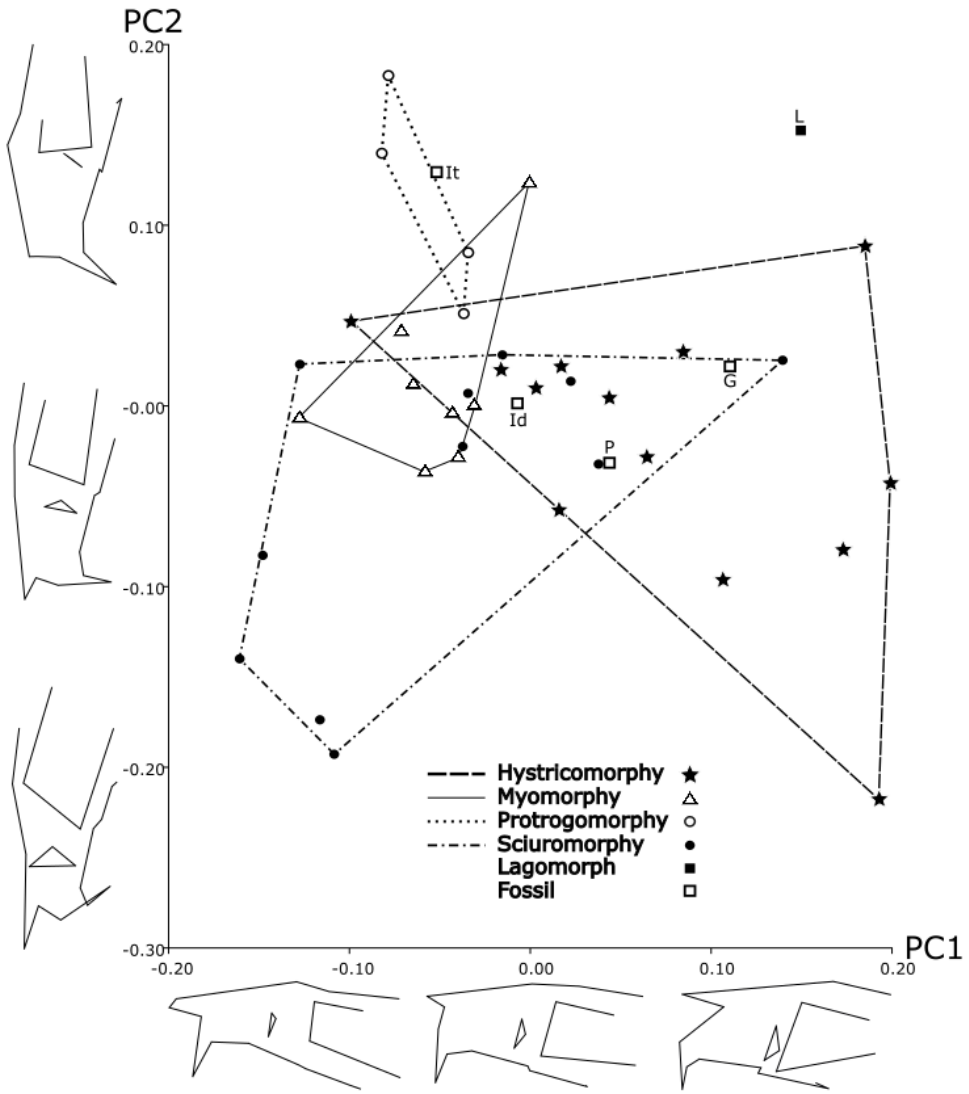


Figure 32: Cranial shape variation across the first two principal components. Boundaries have been superimposed onto extant rodents to indicate the space occupied by members of each anatomical condition of the masseter (inset) with wireframes to indicate general pattern of morphological change along each axis. Id = *Ischyromys douglassi* (USNM 617532), It = *I. typus* (USNM 16828), G = *Gomphos elkema* (MAE-BU 14467), L = *Lepus californicus* (UMZC E3941), P = *Paradelomys crusafonti* (UM ACQ6618).

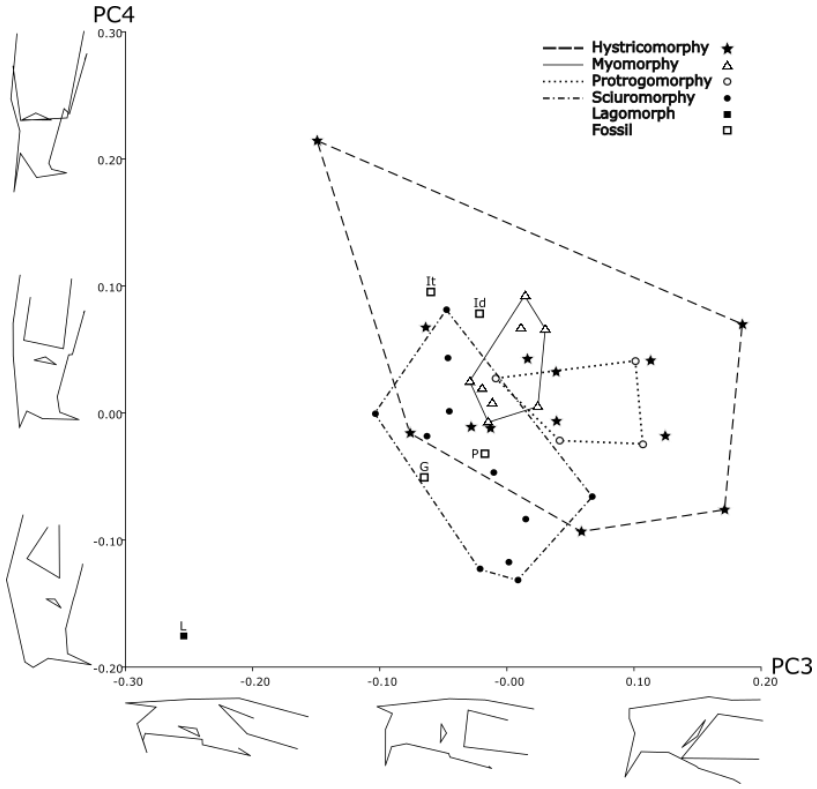


Figure 33: Cranial shape variation across the third and fourth principal components. Boundaries have been superimposed as in Figure 32 to indicate the space occupied by members of each anatomical condition of the masseter (inset). Abbreviations are as in Figure 32.

The plot of the third and fourth principal components (Figure 33) displays more overlap between the different masseter morphologies than PC1 plotted against PC2 (Figure 32) and largely separates rodents from *Lepus*. It also explains less cranial shape variation (21.3%). Anatomically hystricomorphous species occupy a larger cluster in this plot compared to sciuromorphous species. As in the plot of PC1 and PC2, myomorphous and protrogomorphous species occupy the tightest clusters. *Ischyromys douglassi* and *I. typus* are close to each other in morphospace on both PC3 and PC4 and occupy the hystricomorphous cluster. *Gomphos* occupies the sciuromorphous cluster, while *Paradelomys* occupies the area of overlap between the sciuromorphous and hystricomorphous clusters.

Based on the phylogenetic study of Asher et al. (2019), *Ischyromys* is well supported (posterior probability of 0.94) as a member of the squirrel-related clade (Sciuromorpha), sister to *Aplodontia* - Sciuridae (Figure 34). Anatomically, and in agreement with Wood (1976), the genus *Ischyromys* shows polymorphic masseter architecture, with some species (e.g., *I. douglassi*) exhibiting anatomical sciuromorphy and others (e.g., *I. typus*) anatomical protrogomorphy (Figure 30). Figure 34 provides a reconstruction of masseter types as a single, multistate character mapped onto the phylogeny of Asher et al. (2019: figure 2). Several key taxa have not yet been incorporated into that phylogenetic analysis, such as anatomically hystricomorphous theridomyids (Vianey Liaud et al., 2016),

protrogomorphous *Eogliravus* (Storch and Seiffert, 2007), and protrogomorphous, possible stem muroids such as *Knightomys* (Wood, 1965) and *Pauromys* (Dawson, 1968). Nonetheless, optimal topologies from Asher et al. (2019) support the novel interpretation that the anatomically hystricomorphous *Tataromys* comprises the sister taxon of crown Rodentia, and that anatomically polymorphic *Ischyromys* (with sciuromorphous *I. douglassi* and protrogomorphous *I. typus*) comprises the sister taxon of *Aplodontia* plus *sciurids*.

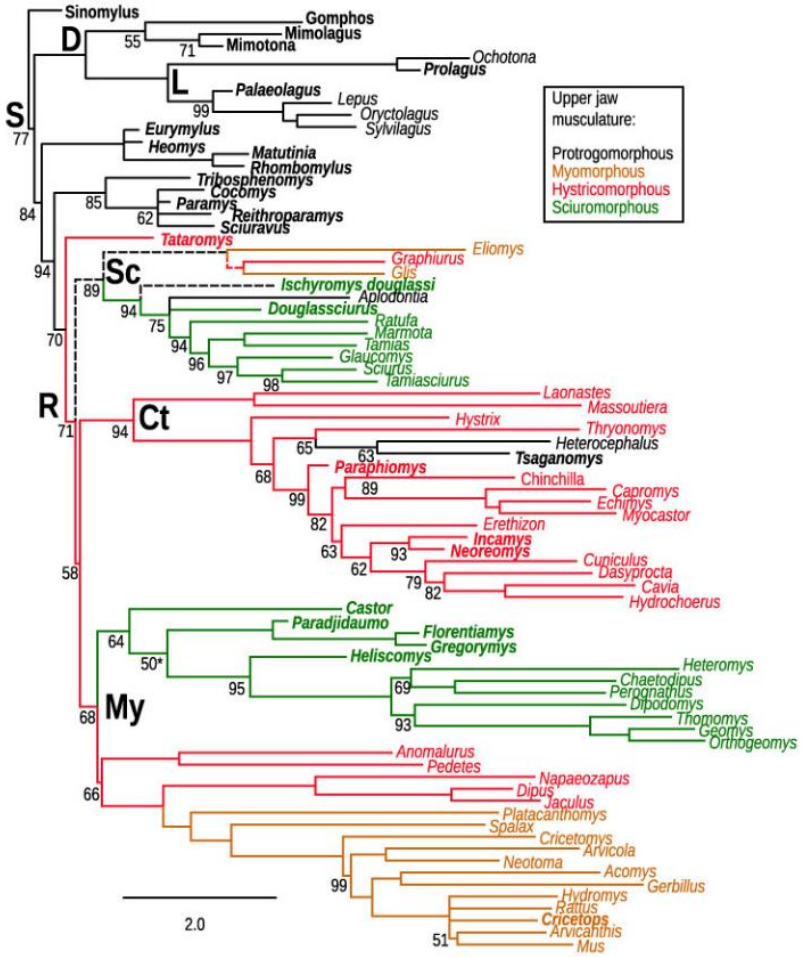


Figure 34: Bayesian phylogenetic estimate from Asher et al. (2019), derived from an alignment of eight nuclear and six mitochondrial genes concatenated with DNA indels and 219 morphological characters, comprised of majority rule consensus of 17500 post-burnin (50% of 35G generations sampling every 1000) topologies with posterior probabilities shown as percentages adjacent to each node. Only total group Glires are shown. Numbers adjacent to internal nodes show majority rule consensus across post-burnin trees (representative of Bayesian posterior probabilities); no number indicates 100%. Bold indicates fossils; branch lengths correspond to scale at bottom. Abbreviations are Ct for Ctenohystrica, D Duplicidentata, Gl Glires, L Lagomorpha, My Myomorpha, R Rodentia, S Simplicidentata, and Sc Sciuromorpha. Anatomically protrogomorphous taxa and parsimony-reconstructed branches are shown in black, myomorphous orange, hystricomorphous red, and sciuromorphous green. Unambiguous state reconstructions are shown with solid lines, ambiguous with dashed. Masticatory categories follow Maier and Schrenk (1987), Wang (1997), Meng et al. (2003), Marivaux et al. (2004), Wible et al. (2005), Hautier et al. (2008), Cox and Faulkes (2016), and this chapter.

Discussion

Among *Ischyromys* teeth from West Canyon Creek, Heaton (1996, p. 395) noted "a unique feature that separates it from all other species of *Ischyromys*... the exceptionally high incidence of medial and lingual accessory cusps." However, such accessory cuspules are not unique to the West Canyon Creek specimens but are also evident in *I. douglassi* from McCarty's Mountain. In his description of *I. douglassi,* Black (1968, p. 286), describes this as "a small cusp on the posterior face of the protoconid which bulges into the basin behind the metalophid" and that "this structure is quite prominent on M1-M3" and it is shown in his figures 10–12. Black's illustrations (1968: figures 10–12) of the lower dentition (CMNH 1053, 1125, and 10963) show more wear than the lower molars of USNM 617532 and other specimens from West Canyon Creek, but they are consistent with the presence of lower accessory cuspules in *I. douglassi*, evident in USNM 475454, 489144 (Figure 28), 617532 (Figure 26) and other specimens from West Canyon Creek.

One of the most striking features of *Ischyromys douglassi* (and in particular USNM 617532) is the presence of a broad zygomatic plate and muscle scars of the deep masseter reaching onto the rostrum, anterodorsal to the infraorbital foramen (Figure 24b, Figure 30b, d). These are not as extensive or well defined as in extant sciuromorphous taxa such as *Sciurus*, *Castor*, or *Ratufa* (Figure 24c) but neither are they present in anatomically protrogomorphous rodents (Figure 24a, Figure 30g–h). In agreement with Wood (1976), species of *Ischyromys* such as *I. douglassi* and *I. veterior* (Wood, 1976) match Brandt's (1855) definition of anatomical sciuromorphy. Previous generations of paleomammalogists have often mixed the anatomical conditions of the masseter with rodent taxonomy, leading to the widespread, historical classification of *Ischyromys* among 'protrogomorph'-grade rodents (Wood, 1937; Chaline and Mein, 1979; Fahlbusch, 1985). Black (1968) was concerned that crushing had given Wood's fossils an altered appearance, but that is not the case here. Although parts of the occiput, braincase, nasals, and premaxilla show damage (Figure 25, Figure 30a), the specimen (and particularly the infraorbital region) of USNM 617532 is otherwise well-preserved (Figure 30b–d) and consistent with our interpretation that one of the oldest specimens of *Ischyromys* exhibits a zygomatic plate that extends anterior to the infraorbital foramen. It is, thus, anatomically sciuromorph.

Anatomical sciuromorphy has long been defined in the literature based on expansion of the deep masseter along the anterior zygomatic arch (Hautier et al., 2008; Cox et al., 2012). The first two principal components of the geometric morphometric analysis undertaken here place USNM 617532 within the shape space occupied by extant, anatomically sciuromorphous rodents and outside of that exhibited by extant protrogomorphous rodents (Figure 32). However, there is substantial overlap among the hystricomorph, sciuromorph, and myomorph clusters based on the first two principal

components (Figure 32), which collectively account for about 31% of the variation, and even more overlap among all categories based on extant species using principal components 3 and 4 (Figure 33, accounting for ca. 21% of the variation). The cranial landmarks used here show that there is a great amount of shared shape variation among these rodents, despite their different masseter structures. Morris et al. (2018) investigated convergences in cranial variation across Euarchontoglires, and their results reflect this finding in rodents. They found that the cranial variation within the squirrel-related clade of rodents, most of which are anatomically sciuromorphous (except for *Aplodontia* and glirids; see Figure 34 and Maier et al., 2002), overlapped with that of both mouse-related and ctenohystrican clades. When considering the first two principal components in our study (Figure 32), *Ischyromys douglassi* overlaps with the sciuromorphous and hystricomorphous clusters, but not the protrogomorphous cluster (as defined by extant rodents only). In contrast, *I. typus* appears adjacent to the protrogomorph cluster. Their distance in morphospace further demonstrates that although these two species are closely related, their rostral morphology (and by extension myology) is different. When the third and fourth principal components are considered (Figure 33), both species of *Ischyromys* appear within the hystricomorphous cluster and close to the margins of the sciuromorphous and myomorphous clusters.

The landmarks chosen for this analysis likely capture some shape variation not directly related to the masseter morphology, and each principal component captures a slightly different aspect of variation. Thus, *Ischyromys typus* may not resemble other protrogomorphous species closely due to other aspects of the skull shape that are not shared, or they might be represented by the other principal components. The species we regard as anatomically protrogomorphous are all visually quite different. For example, *Heterocephalus glaber* has relatively smaller infraorbital foramina than *I. typus* and *Aplodontia rufa.* Furthermore, although the masseter in each of these species never penetrates the infraorbital foramen, it comes close in some bathyergoids (Cox and Faulkes, 2014). The zygomaticomandibularis of *H. glaber* extends past the zygoma until it reaches the anterior orbit, but does not pass through the small infraorbital foramen (Cox and Faulkes, 2014). Although a member of Ctenohystrica (an otherwise exclusively hystricomorphous group), bathyergoids have a secondarily derived protrogomorphous condition, which is likely to be an adaptation for fossorial life (Cox and Faulkes, 2014).

Much like the other masseteric conditions, protrogomorphy displays homoplasy, having evolved multiple times among the different rodent clades. Following the phylogenetic study of Asher et al. (2019), the anatomically protrogomorphous *Heterocephalus* and *Tsaganomys* are both deeply nested within the clade of Ctenohystrica and share multiple common ancestors with anatomically hystricomorphous species such as *Thryonomys*, *Hystrix*, and ctenodactylids. Similarly,

protrogomorphous *Aplodontia* is nested within the squirrel-related clade and shares common ancestors with sciuromorphous *Ischyromys douglassi* (and by extension the protrogomorphous *I. typus*), and myo- and hystricomorphous glirids (Figure 34).

Albeit weakly supported (posterior probability of 71 in Figure 34), the node separating crown Rodentia from the anatomically hystricomorphous taxon *Tataromys* (from the Oligocene-Miocene of central Asia; Wang, 1997) underscores the possibility that the ancestor of crown rodents was itself non-protrogomorphous. *Tataromys* has generally been regarded as an Asian relative of ctenodactylids (Wang, 1997; Marivaux et al., 2004; Wible et al., 2005; Oliver and Daxner-Höck, 2017), not a simplicidentate outside of crown Rodentia. However, *Tataromys* was reconstructed using both probabilistic optimality criteria and parsimony (Asher et al., 2019: figure 4 and S2) as the sister taxon of crown rodents, not a member of Ctenohystrica (Figure 34). Further phylogenetic analyses with better samples of (for example) fossil Ctenohystrica would be necessary to further test this possibility.

Swanson et al. (2019) considered the evolution of rodent masseteric architecture in a phylogenetic study of a large UCE dataset. They discussed several key fossils, including anatomically protrogomorphous species that, according to the literature (Wood, 1959, 1965, 1985; Dawson, 1968; Marivaux et al., 2004; Storch and Seiffert, 2007), are likely stem glirids (*Eogliravus*), stem mouse-related (*Pauromys, Prolapsus, Knightomys*), or sister to crown Rodentia (*Bumbanomys*). With these fossils in their literature-inferred placements, Swanson et al. (2019: figure 2) reconstructed the common ancestors of two of the three major rodent clades (excluding Ctenohystrica), and of Rodentia itself, as most likely anatomically protrogomorph, consistent with the long-held and legitimate inference that, at some point, the total clade Simplicidentata (encompassing Rodentia) evolved from an anatomically protrogomorph common ancestor.

The phylogenetic relationships of the 51 extant taxa sampled by Swanson et al. (2019) were based on a large molecular dataset, comprising an alignment over 2000 UCEs (ultraconserved elements, or genomic fragments of DNA) summing to nearly 900 kb. Asher et al. (2019) sampled a much smaller molecular dataset (ca. 17kb) for slightly more extant taxa (60), plus fossilizable morphological characters sampled for all living taxa and another 42 fossils. These two datasets yielded highly congruent topologies overall, but with differences in the position of the rodent root. Swanson et al. (2019) reconstructed the mouse related clade (Myomorpha) as sister to Ctenohystrica-Sciuromorpha; Asher et al. (2019; shown here in Figure 34) reconstructed Sciuromorpha as sister to Ctenohystrica-Myomorpha. The much larger sequence dataset of Swanson et al. (2019) comprises a stronger basis on which to hypothesise the affinities of extant rodents. However, their fossils were positioned in their tree based on the literature, not on a phylogenetic data matrix including those fossils along with extant

species. In contrast, Asher et al. (2019) did undertake a phylogenetic analysis of DNA and morphology for living taxa along with morphological data for fossils. In order to definitively arbitrate between the competing hypotheses regarding masseteric architectures at key nodes in the early radiation of rodents, it will clearly be necessary to integrate the samples of living taxa and fossils discussed by both Swanson et al. (2019) and Asher et al. (2019), for example by including anatomically protrogomorphous (e.g., *Eogliravus*) and non-protrogomorphous (e.g., *Tataromys*) fossils long assumed to be within crown Rodentia, but seldom tested in a total evidence framework.

Heaton's (1996) cluster analysis of 31 ischyromyid populations, based on dental and jaw morphometrics, shown in his figure 14 and without the data on rostrum morphology discussed here, provides evidence of further masseteric homoplasy within ischyromyids. Populations he describes as having the sciuromorphous condition, placed in *Ischyromys* (*Titanotheriomys*), form a cluster with two populations, West Canyon Creek and Porvenir (late and latest Duchesnean of Wyoming and Texas, respectively). These are geologically older than the other *Ischyromys* populations in his sample. Heaton (1996) considered the West Canyon Creek population to be a new species, but still categorized it as anatomically protrogomorph (Heaton, 1996: table 8), despite clustering with Wood's sciuromorphous *Titanotheriomys*. He suggests later on in a proposed evolutionary tree (Heaton, 1996: figure 15), that the West Canyon Creek population is the basal-most species of *Ischyromys* and that (based on its "*Ischyromys* like skull" and "lack of an elongate m3") it "makes a likely ancestor for all later species" (1996, p. 395). There are few populations of *Ischyromys* that predate West Canyon Creek (Lac Pelletier, Badwater Creek, and possibly Porvenir), but their lack of cranial and associated material have so far made it impossible to identify species or determine the anatomical features of the masseter (Heaton, 1996). Thus, *I. douglassi* from West Canyon Creek (including USNM 617532), is the oldest, anatomically best-known ischyromyine, and to our knowledge it is also the oldest rodent known to be anatomically sciuromorph.

Conclusion

Rodent classification now reflects the structure of the increasingly well-corroborated phylogenetic tree (Marivaux et al., 2004; Churakov et al., 2010; Fabre et al., 2012; Asher et al., 2019; Swanson et al., 2019), not typological concepts of rodent masticatory patterns as articulated in the nineteeth century and used to varying extents well into the twentieth century. Authors such as Wood (1965), Luckett and Hartenberger (1985), Korth (1994) and many others were aware of the homoplasy of rodent masticatory types of the upper jaw. However, most authors prior to this decade did not have

the advantage of the well-corroborated evolutionary tree upon which to anchor their taxonomies or reconstruct character evolution. We, therefore, use this phylogeny to recognise the paraphyly of 'ischyromyoids', some of which (e.g., *Paramys, Reithroparamys*) are outside of crown Rodentia and some of which (e.g., *Ischryomys*) are within the squirrel-related clade (or Sciuromorpha, Figure 34). Our results also underscore the long-recognised homoplasy in the masticatory musculature of the upper jaw, including within the genus *Ischyromys*.

Previous authors have often referred to ischyromyine rodents as 'protrogomorphous', in some cases implying the anatomical condition (Heaton, 1996), others a taxonomic grade (Dawson, 1977; Vianey-Liaud, 1985; Korth, 1994) and sometimes a mixture of both (Wood, 1980). Anatomically, our study confirms the views of Wood (1976) that at least some members of this group, including *Ischyromys douglassi,* possess diagnostic features of sciuromorphy. Geometric morphometrics demonstrate cranial similarities between this species and extant rodents, whilst highlighting the mosaic nature of jaw musculature evolution across rodents. The age of this specimen, and phylogenetic affinities postulated by Heaton (1996: figure 15), suggest that rather than being the derived condition, an anatomically sciuromorphous rostrum characterizes the geologically oldest populations of *Ischyromys* yet known. Our phylogenetic placement of the fossil *Tataromys* as a non-rodent simplicidentate, and the possible, non-protrogomorphous character optimizations at the common ancestors of major rodent clades, and even for Rodentia itself (Figure 34), is tentative pending a phylogenetic analysis that includes more fossils. Nonetheless, our results underscore the possibility that one or more early-diverging rodent clades were not anatomically protrogomorph.

Chapter Six: general discussion and concluding remarks

Muscardinus avellanarius

Exploring layers of the Earth's strata reveals a menagerie of lifeforms and ecosystems, which are no longer in existence, as well as a sequential record of how this diversity has changed over time. Fossil specimens represent a window into the deep past and they are necessary to form an understanding of the evolution of life on Earth. Throughout this thesis are examples of how the information content of fossils can be used to infer evolutionary history, from tracing the movement of taxa across the globe to understanding when complex adaptations first evolved. The story of how Glires arrived at their current success spans over 60 million years' worth of fossil material as well as increasingly large amounts of molecular data. The aim of this thesis has been to use this evidence to gain a deeper understanding of Glires origins and diversification, as well as to provide critique on the efficacy of fossil data itself.

The importance of Asia and total-evidence analyses

The origins of many mammal groups, including Glires, are thought to have occurred on the Asian continent, due to the abundance of early stem taxa from that part of the world (Dawson and Beard, 1996; Beard, 1998). Results from Chapter Four of this thesis add further evidence to support the importance of Asia as the 'cradle' of Glires evolution. My total-evidence analyses of fossil and molecular data point towards a Cretaceous origin of Glires in Asia before they diversified and spread across the world. However, the question of where each crown rodent group originated and which of these three clades is positioned at the rodent root could not be resolved by my analyses. This is in part due to my addition of an emerging type of molecular data, ultraconserved elements (UCEs), which although congruent with other molecular data regarding many relationships within Rodentia, are also known to produce a different relationship at the rodent root (Swanson et al., 2019). My morphology and protein coding-based phylogeny proposes that Sciuromorpha is at the root of rodents, whereas the dataset which includes UCEs as well as the morphology and protein coding DNA implies that Rodentia is rooted by Myomorpha. Conflicting hypotheses at the rodent root are common (Huchon et al., 2007; Fabre et al. 2012; Swanson et al, 2019) and have resulted in the relationship being represented as an unresolved trichotomy in summaries of rodent systematics (Elia et al., 2019).

The placement of key fossil taxa also differed between my two total-evidence phylogenies, resulting in different interpretations of common ancestors. For example, the position of the Asian fossil *Tataromys*, either implies that Ctenohystrica + Myomorpha had a ctenodactyloid-like common ancestor (morph+coding phylogeny) or that Rodentia itself descended from an ancestor with ctenodactyloid-like features (UCE phylogeny). My results also show that topological differences inferred by different total-evidence phylogenies can result in alternative theories of historical

biogeography. A key finding was that both Sciuromorpha and Myomorpha were found to have North American origins according to the morph+coding phylogeny. This result is reminiscent of the 'Paleocene paradox', whereby the earliest members of a group appear in North America while their presumed ancestors were still roaming in Asia. In contrast, the UCE-based phylogeny places Sciuromorpha and Myomorpha anywhere but North America during their initial diversification. This uncertainty opens up the possibility that North America may have had a more nuanced role in the early evolution of Rodentia than first thought. Beard's (1998) proposition that North America was a 'cul de sac' of Asia could therefore, be oversimplified.

Future work in this area would benefit from the discovery of new Paleocene and Eocene fossil material from both sides of the Pacific, to increase the resolution of data surrounding the 'Paleocene paradox' in Rodentia. Fortunately, new data is always coming to the fore with the Erlian Basin of northern China proving particularly productive. The Erlian Basin covers the Gashatan, Bumbanian, Arshantan, and Irdinmanhan Asian land mammal ages (Paleocene to the middle Eocene; Meng et al., 2007), making it a good place to look for clues concerning the Paleocene paradox. In the past year alone (2020-2021), two tsaganomyids (Li, 2021), six cylindrodontids (Li, 2020; Li, 2021), three ischyromyids (Fostowicz-Frelik et al., 2021) and two ctenodactyloids (Xu and Li, 2020; Li, 2021) have been described from the Erlian Basin. The recent description of new 'ischyromyid' material from the Eocene of the Erlian Basin is of interest because until recently, ischyromyids were regarded as rarities in Central Asia (Fostowicz-Frelik et al., 2021). This new material shows that ischyromyids had a much bigger presence on the Asian continent than first thought, although their age suggests that ischyromyids did not originate in Asia (Fostowicz-Frelik et al., 2021). Asian ischyromyids are still less diverse and a bit younger than their Paleocene cousins in North America, strengthening the idea that movement of some rodent groups was *into Asia* rather than from it. Further work on material from the Erlian Basin will help to clarify the movement and distribution of rodents shared between the two continents, as will further studies on the correlation between the Gashatan-Bumbanian and Paleocene-Eocene boundaries.

Furthermore, the contrasting results between my analyses which include UCEs and those that do not, highlight the consequences of data selection on phylogeny. Little work has combined fossil data with UCEs in total-evidence analyses (Chen et al., 2019; Scarpetta, 2020), but my results demonstrate that joining fossil data with large genomic partitions can alter the phylogenetic placement of fossils. Combining fossil data and UCEs however, does result in large amounts of missing data. Simulation and empirical studies by Wiens (Wiens, 1998; Wiens, 2003; Wiens et al., 2010) indicate that large amounts of randomly distributed missing data in phylogenetic analyses are not problematic for phylogenetic reconstruction, but there is increasing evidence to suggest that non-randomly distributed missing data can produce topological artefacts when probabilistic methods, such as Maximum Likelihood, are used

(Simmons, 2011; Simmons, 2014). Furthermore, Scarpetta (2020) found that the signal from UCEs might overpower that of morphological data. Nevertheless, the benefits of combining data under a total-evidence framework (Gatesy and Arctander, 2000; Gatesy and Baker, 2005) merit further integration of fossil and UCE data. Analyses that incorporate UCEs are already computationally expensive, but including complex, morphological matrices with missing data requires additional models and computational power. Bayesian methods, which allow tip-dating, complex models and various partition schemes, would be an ideal choice for analysing UCE and fossil data together, but computational constraints would prevent this approach for most datasets. Yet the type of phylogenetic method used to analyse UCE data can have significant consequences for topology, more so than alignment or trimming methods (Portik and Wiens, 2021). Since future work will likely create more genomic data, not less, we must prioritise the development of programmes that can handle the computational challenges of combining fossil and UCE data under the most appropriate probabilistic frameworks. It is worth noting that morphometric data has already been successfully combined with UCE data in total-evidence analyses of species delimitation (Prebus, 2021). Another avenue of research could therefore explore combining morphometric PCA data derived from fossil specimens with UCEs in a phylogenetic context.

The Tibetan Plateau as a refuge for Paleogene Glires

As well as tracing the biogeographic roots of Glires to Asia, my thesis also investigated whether the formation of the Tibetan Plateau in Central Asia had an impact on the diversification of Glires. My results from Chapter Four indicate that initial diversification of rodent crown groups coincided with the collision of the Indian plate with Southeast Asia and the Paleocene-Eocene Thermal Maximum. It is hypothesised that harsh environmental conditions, and topographic complexity caused by severe tectonic activity, will lead to increased rates of speciation (Badgley, 2010; Fortelius et al., 2014). The burst of diversification in Rodentia during this time seems to conform to this theory. My results also observe instances where drops in rodent diversity during warm periods are followed by diversification of Asian taxa during subsequent cold snaps. For example, it appears that the diversity of Glires dipped during the Mid Eocene Thermal Maximum and Late Oligocene Warming. The latter event was proceeded by the diversification of dipodoids in Asia during the Oligocene glaciations, and the former was followed by the appearance of Asian muroids and cricetids during the Miocene glaciations. There is now substantial evidence to suggest that the Tibetan Plateau acted as a refuge for cold adapted taxa during the Pliocene, and that these animals then diversified and expanded their range during the Pleistocene ‘Ice Age’, when global temperatures plummeted (Tseng et al., 2013a; Tseng et al., 2013b; Deng et al., 2015; Wang et al., 2015). This ‘out of Tibet’ theory (Deng et al., 2015) could explain the

patterns observed in my analyses of Glires diversification and suggest that the Tibetan Plateau might have acted as a refuge for Glires during the Paleogene. However, studies of Pliocene and Pleistocene fauna have the advantage of better sampling of small mammals than that of the Oligocene or Miocene glaciations. Further additions to the fossil record of Glires from Central Asia during these earlier times would be needed to confirm the role of the Tibetan Plateau as a refuge for cold-adapted taxa during the Paleogene. Another source of evidence could potentially come from extant species themselves. A recent study by Wang et al. (2020) combined pika phylogeography with whole-genome sequencing and found that genes conferring cold-tolerance are associated with the earliest pikas, whose origins lie in the Tibetan Plateau. It would be interesting to carry out similar studies on other central Asian taxa, such as zokors, cricetids and dipodoids, to see if their genomic evidence also points towards the Tibetan Plateau acting as a 'training ground' for cold-tolerant taxa.

As our climate increases in temperature due to global warming, the role of the world's third largest store of ice is likely to have a big impact on the future of Glires. Further work might model the distribution changes of small mammals in response to predicted temperature increases, to investigate whether the Tibetan Plateau might serve as a refuge for cold-adapted Glires in the coming years. Several studies have modelled distribution changes of central Asian plants (Yu et al., 2018; Xiong et al., 2019) and ungulates (Wu et al., 2017; Zhang et al., 2021) in response to different climate change scenarios, with habitat reduction, fragmentation and migration to higher altitude areas all found to be common trends. However, no work has attempted to predict if small mammals might seek refuge in the Tibetan Plateau under expected climate changes over the next 50 to 100 years. Ecological niche models and climate models, combined with knowledge of Quaternary refugia, could be used to map the expected distribution of Glires over the coming years and importantly, help direct conservation efforts in Asia.

Derived musculature in Eocene fossil Glires

One of the benefits of using fossil data in phylogenetic analyses is that extinct taxa often have a different combination of character states to living species, and sometimes they may even possess features not seen in living taxa at all (Koch and Parry, 2020). In combination with phylogeny, the morphology of fossils can change our ideas of character homology and inform us about how adaptations might have evolved over time. However, the interpretation of fossil anatomy is not always straightforward and there may be disagreements between researchers on how to code particular characters. In Chapter Five of this thesis, I explore the jaw musculature of *Ischyromys* species from the Eocene. *Ischyromys* belongs to the paraphyletic group 'Ischyromyidae', which is widely considered to

be anatomically protrogomorphous (Korth, 1994). However, my observations of a well-preserved specimen of *Ischyromys douglassi* find evidence of a zygomatic plate and scars of the masseter muscle reaching onto the rostrum, both typical of anatomical sciuromorphy. This finding supports Wood's (1976) belief that some 'ischyromyids' were sciuromorphous, which was heavily criticised at the time by others, specifically Black (1968), who had examined the same specimens as Wood. Furthermore, my geometric morphometric analyses indicate that the cranium of *I. douglassi* resembles living taxa more than other members of *Ischyromys*. This, coupled with the age of the specimen (Duchensean), implies that protrogomorphy is not the ancestral state of *Ischyromys,* and that sciuromorphy is. In fact, the possible position of *Tataromys* (an anatomically hystricomorphous species) at the root of Rodentia (Chapter Four, Figure 20; also Asher et al., 2019 and Rankin et al., 2020), could indicate that protrogomorphy is not the ancestral condition of Rodentia as a whole either.

The identification of specimen USNM 617532 as *Ischyromys douglassi*, and the reassignment of this species from anatomically protrogomorphous to sciuromorphous, highlights the benefits of revisiting museum collections and previous debates on anatomical structure. The homoplasy of masseter types amongst rodents has been observed time and time again (Luckett and Hartenberger, 1985; Korth, 1994; Samson et al., 2019). It seems likely that more specimens of extinct Glires that were initially identified as 'protrogomorphous', could upon further investigation, be found to have more complex masseter architectures. As well as the general need to gather new fossil Glires from the Eocene and Paleocene, it would be beneficial to revisit material already held in collections. Methods that were not available during the 19th and early 20th centuries, such as microCT and geometric morphometrics, should be used to provide new detail on specimens that are rarely seen.

Ischyromys is well represented in museum collections, but as of yet, no phylogenetic study has attempted to include all of the members of this genus. A potential overhaul of *Ischyromys* systematics could be achieved and is desirable, given their importance in mapping the evolutionary history of masseter musculature in Glires. Furthermore, the uncertain relationship between *Ischyromys* and '*Titanotheriomys*' has not been revisited in depth since Heaton's (1996) cluster analyses of dental data. Future work could undertake a phylogenetic analysis that incorporates material from all known species of *Ischyromys* (including material Wood referred to as '*Titanotheriomys*'), of which there are seven currently recognised (Korth, 1994). In addition, the anatomical musculature of these species should be investigated more fully. As shown in Chapter Five of this thesis, anatomical protrogomorphy may not be the ancestral condition in *Ischyromys* and it is possible that species other than *I. douglassi* may have had different masseter arrangements. The anatomical sciuromorphy observed in *I. douglassi* is not identical to that seen in modern day Glires, and as well as intermediary stages, we might tentatively consider the possibility that masseter arrangements could exist that are not represented

by any extant rodents. To build upon the geometric morphometric analysis of 3D landmarks carried out in Chapter Five, further studies would likely benefit from the application of semi-sliding landmarks around the zygomatic arch and infraorbital foramen. By using semi-landmarks, the resolution of cranial phenotypic disparity between species of *Ischyromys* could be increased. Semi-landmarks have been used before by Hautier et al. (2012) to report on the morphological variation present in the infraorbital region of hystricomorphous rodents. By using this method, they were able to show that variation in infraorbital shape was largely explained by phylogeny and that 'hystricomorphy' actually consists of several morphotypes. The implications of this are that infraorbital foramen shape could potentially aid the phylogenetic placement of rodents (Hautier et al., 2012). Semi-sliding landmarks would likely clarify the ambiguous zygomasseteric structure of some fossils such as '*Titanotheriomys*' and may even help to resolve their phylogenetic relationships, but it may be necessary to consider that some muscle arrangements may be novel and not comparable to those seen today.

The information of fossil Glires

Discussions of missing data and the efficacy of fossil remains in phylogenetic analysis often go hand in hand. Since the cladistics revolution in the late 1970s, there has been a large amount of work testing what effect fossils and missing data have on reconstructing phylogeny (Patterson, 1981; Hennig et al., 1981; Gauthier et al., 1988; Donoghue et al., 1989; Wiens, 1998; Wiens, 2003a). More recently, studies which seek to artificially 'fossilize' living taxa of known phylogenetic affinity have been used to test whether or not fossils might distort phylogeny (Asher and Hofreiter, 2006; Springer et al., 2007; Sansom and Wills, 2013; Pattinson et al., 2015). These experiments however, have not reached a consensus. For example, Sansom and Wills (2013) found that removing soft tissue characters of living species led to a reduction in congruence with their optimal tree, as well as taxa appearing in more stemward positions than they were thought to occupy. In contrast, Pattinson et al. (2015), who based their artificial fossils on real fossil templates, found that fossil Primates were capable of accurately constructing phylogeny. In Chapter Two of this thesis, I employ the same artificial extinction methods as Pattinson et al. (2015) to fossil Glires. My results indicate that fossil Glires contain phylogenetic signal capable of accurately reconstructing known phylogeny and can do so even at very high levels of missing data. Furthermore, my results also found that there are no biases towards stemward or crownward slippage when fossil Glires are included in phylogenetic analyses, and so there is no bias towards producing misleadingly primitive trees. Both of these results are very encouraging within a field where missing data are ubiquitous.

Another aspect of missing data that this thesis aimed to address is the nuanced effects of partition bias. It is known that different morphological partitions can contain different phylogenetic signal to one another (Pattinson et al., 2015; Mounce et al., 2016), and it has been proposed that some partitions may be more congruent with molecular based hypotheses of phylogeny than others (Sansom et al., 2017). In Chapter Three of this thesis, I investigate another facet of phylogenetic signal, namely whether or not the phylogenetic signal contained within morphological partitions is localised to particular parts of the tree. I find that this is indeed the case, with partitions such as dental data optimizing as apomorphies closer to the root of the tree and cranial partitions optimizing as apomorphies closer to the tips of the tree. I presumed that this might have a large impact on phylogenetic analyses that contain fossil data, as extinct taxa are usually over-represented by only a few partitions. Furthermore, results from Pattinson et al. (2015) and from Chapter Two in this thesis indicate that sampling across partitions could be beneficial to accurate phylogenetic reconstruction. I hypothesised that removing cranial data, which appears to define more shallow clades, would result in stemward slippage. Small mammal fossils in particular often lack cranial and postcranial data, and many comprise of solely dental data. It is not uncommon for phylogenetic analyses to be based almost entirely on dentition (Heaton, 1996; Marivaux et al., 2004; Rodrigues et al., 2009; Antoine et al., 2011). However, I found that removing localised phylogenetic signal did not result in the expected patterns of slippage. This is in part due to 'hidden signal' shared by multiple morphological partitions. The other reason is that partitions with localised signal also contain characters which optimize at different depths of the tree. For dental data in particular, while it may define more basal clades, it also contains key characters which optimize on more shallow branches of the tree. It is encouraging that the morphological characters used in research capture phylogenetic signal across a wide portion of rodent evolutionary history, within each morphological partition. It is still good practice of course, as highlighted by the work in this thesis, to sample across morphological partitions in order to reap the full benefits of hidden signal.

The artificial extinction and morphological partition experiments in this study have been very informative, but these methods could be expanded upon to widen their scope and relevance. For example, it would be interesting to incorporate alternative measures of topological congruence with which to test morphological datasets, such as stratigraphic congruence. Stratigraphic congruence relies on the straightforward assumption that branching events in a phylogeny should reflect the order of appearance of taxa in the strata. Like fossil data, the geological record is itself incomplete and is subject to preservational bias, however, it is also independent of phylogeny. Instead of measuring the congruence of experimental topologies against a well-corroborated tree, which implies some prior knowledge of the 'true' tree, topologies could be directly assessed against the stratigraphic record.

This approach proved successful in Sansom et al's (2018) meta-analysis of vertebrate morphological datasets, which was used to assess the advantages of maximum parsimony over probabilistic methods. The method of phylogenetic analysis performed in artificial extinction and morphological partition experiments is another area that could be revisited. Artificial experiments require the reconstruction of many thousands of topologies and so phylogenetic methods that are less computationally intensive, such as maximum parsimony, are intuitive. However, it is currently debated whether or not maximum parsimony can outperform probabilistic methods when it comes to analysing morphological data and missing data. For example, Sansom et al's (2018) study found that Bayesian topologies were less congruent with stratigraphic data, than maximum parsimony topologies, but more recently, a large simulation study by Koch et al. (2021) suggested that probabilistic methods produce more congruent topologies than maximum parsimony. Furthermore, Koch et al. (2021) found that the advantage conferred by probabilistic methods becomes more significant as the amount of missing data increases. Given this finding, future artificial extinction and morphological partition experiments should explore Bayesian and Maximum Likelihood methods when testing the ability of fossils to accurately reconstruct phylogeny. Incorporating probabilistic phylogenetic analysis into artificial extinction experiments would make this type of research increasingly relevant in a field where combining fossil and molecular data under a probabilistic, total-evidence framework is becoming more common.

Summary

This thesis aims to examine the efficacy of fossil Glires remains in phylogenetic analysis, and then to use this fossil material to explore the Asian origin of Glires, as well as the evolution of their incredibly effective jaw musculature. Through artificial extinction experiments and the exploration of phylogenetic signal within morphological partitions, I find that fossil Glires, despite their missing data and partition biases, are capable of accurately reconstructing phylogeny. I also find evidence for localised phylogenetic signal within morphological partitions. My total-evidence phylogenetic analyses support an Asian origin of Glires and indicate that the Tibetan Plateau might have had a strong influence on the initial diversification Rodentia, as well as acting as a possible refuge for cold-adapted rodents during the Oligocene and Miocene glaciations. However, my results also highlight conflict between the role of Asia and North America in the diversification of Myomorpha and Sciuromorpha. Close examination of the North American fossil *Ischyromys douglassi* indicates that anatomical sciuromorphy, not protrogomorphy, may be the ancestral condition for 'Ischyromyidae'. This emphasises the mosaic evolution of one of the most successful anatomical adaptations possessed by rodents.

References

Allison PA, 1988. Konservat-Lagerstätten: cause and classification. *Paleobiology*, 14, 331–334.

Alroy J, 2017. Effects of habitat disturbance on tropical forest biodiversity. *Proceedings of the National Academy of Sciences USA,* 114, 6056–6061.

Alroy J, 2018. Limits to species richness in terrestrial communities. *Ecology Letters*, 21, 1781–1789.

Altekar G, Dwarkadas S, Huelsenbeck JP and Ronquist F, 2004. Parallel Metropolis-coupled Markov chain Monte Carlo for Bayesian phylogenetic inference. *Bioinformatics,* 20, 407–415.

Anderson D, 2008. Ischyromyidae, 311–325. In: Janis CM, Gunnell GF, Uhen MD (eds), *Evolution of tertiary mammals in North America. Small mammals, xenarthrans, and marine mammals, volume 2*. Cambridge University Press: Cambridge.

Antoine P-O, Marivaux L, Croft DA, Billet G, Ganerød M, Jaramillo C, Martin T, Orliac MJ, Tejada J, Altamirano AJ, Duranthon F, Fanjat G, Rousse S and Salas-Gismondi R, 2012. Middle Eocene rodents from Peruvian Amazonia reveal pattern and timing of caviomorph origins and biogeography. *Proceedings of the Royal Society of London, B*, 279, 1319–1326.

Argand E, 1922. La tectonique de L'Asie. Conférence faite á Bruxelles, le 10 août 1922. *Congrès géologique international (XIIIe session)- Belgique 1922*, 171–372.

Asher RJ and Hofreiter M, 2007. Tenrec phylogeny and the noninvasive extraction of nuclear DNA. *Systematic Biology*, 55, 181–194.

Asher RJ, Meng J, Wible JR, McKenna MC, Rougier GW, Dashzeveg D and Novacek MJ, 2005. Stem Lagomorpha and the antiquity of Glires. *Science*, 307(5712), 1091–1094.

Asher RJ, Geisler JH and Sánchez-Villagra MR, 2008. Morphology, paleontology, and placental mammal phylogeny. *Systematic Biology*, 57(2), 311–317.

Asher RJ, Smith MR, Rankin AH, Emry RJ, 2019. Congruence, fossils and the evolutionary tree of rodents and lagomorphs. *Royal Society Open Science*, 6, 190387.

Ax P, 1897. *The Phylogenetic System*. John Wiley, Chichester.

Bandelt HJ and Dress A, 1986. Reconstructing the shape of a tree from observed similarity data. *Advances in Applied Mathematics*, 7(3), 309–343.

Beard KC, 1998. East of Eden: Asia as an important center of taxonomic origination in mammalian evolution. *Bulletin of Carnegie Museum of Natural History*, 34, 5–39.

Beard KC and Dawson MR, eds., 1998. *Dawn of the age of mammals in Asia*. Bulletin of Carnegie Museum of Natural History: Pittsburgh.

Benton MJ, Donoghue PCJ, Asher RJ, Friedman M, Near TJ and Vinther J, 2015. Constraints on the timescale of animal evolutionary history. *Palaeontologia Electronica,* 18.1.1FC, 1–106.

Bertrand O and Silcox M, 2016. First virtual endocast of a fossil rodent: *Ischyromys typus* (Ischyromyidae, Oligocene) and brain evolution in rodents. *Journal of Vertebrate Paleontology*, 36(3), e1095762.

Black CC, 1968. The Oligocene rodent *Ischyromys* and discussion of the family Ischyromyidae. *Annals of the Carnegie Museum,* 39, 273–305.

Blaimer BB, Ward PS, Schultz TR, Fisher BL and Brady SG, 2018. Paleotropical diversification dominates the evolution of the hyperdiverse ant tribe Crematogastrini (Hymenoptera: Formicidae). *Insect Systematics and Diversity,* 2(5), 3.

Blanga-Kanfi S, Miranda H, Penn O, Pupko T, DeBry RW and Huchon D, 2009. Rodent phylogeny revised: analysis of six nuclear genes from all major rodent clades. *BMC Evolutionary Biology,* 9, 71.

Boivin M, 2017. Rongeurs paléogènes d'Amazonie péruvienne: anatomie, systématique, phylogénie et paléobiogéographie. Unpublished D. Phil Thesis, *Université de Montpellier*.

Boivin M, Marivaux L and Antoine P-O, 2019. New insight from the Paleogene record of Amazonia into the early diversification of Caviomorpha (Hystricognathi, Rodentia): phylogenetic, macroevolutionary and paleobiogeographic implications. *Geodiversitas*, 41, 143–245.

Boos WR and Kuang, 2013. Sensitivity of the South Asian monsoon to elevated and non-elevated heating. *Scientific Reports,* 3(1), 1–4.

Borowiec ML, Rabeling C, Brady SG, Fisher BL, Schultz TR and Ward PS, 2019. Compositional heterogeneity and outgroup choice influence the internal phylogeny of the ants. *Molecular Phylogenetics and Evolution*, 134, 111–121.

Bowen GJ, Clyde PL, Koch PL, Ting SY, Alroy J, Tsubamoto T, Wang YQ and Wang Y, 2002. Mammalian dispersal at the Paleocene-Eocene boundary. *Science*, 295, 2062–2065.

Brady P and Mark MS, 2020. Pseudoextinction analyses and ancestral state reconstruction: evaluating fossil placement accuracy in mammalian phylogenetics. *Society of Vertebrate Paleontology, Annual Meeting 2020*. Poster presentation.

Brandt JF, 1855. *Beiträge zur nähern Kenntniss der Säugethiere Russlands*. Buchdruckerei der Kaiserlichen Akad. der Wissenschaften: St. Petersburg.

Brikiatis L, 2014. The De Geer, Thulean and Beringia routes: key concepts for understanding early Cenozoic biogeography. *Journal of Biogeography,* 41, 1036–1054.

Cappellini E, Welker F, Pandolfi L, Ramos-Madrigal J, Samodova D, Rüther PL, Fotakis AK, Lyon D, Moreno-Mayar JV, Bukhsianidze M and Jersie-Christensen RR, 2019. Early Pleistocene enamel proteome from Dmanisi resolves *Stephanorhinus* phylogeny. *Nature*. 574, 103–107.

Catzeflis FM, Aguilar JP and Jaeger JJ, 1992. Muroid rodents: phylogeny and evolution. *Trends in Ecology & Evolution*, *7*(4), 122–126.

Chaline J and Mein P, 1979. *Les Rongeurs et L'evolution.* Doin: Paris. Charles AJ, Condon DJ, Harding IC, Pälike H, Marshall JEA, Cui Y, Kump L and Croudace IW, 2011. Constraints on the numerical age of the Paleocene-Eocene boundary. *Geochemistry, Geophysics, Geosystems*, 12(6), Q0AA17.

Chatterjee S, Scotese CR and Bajpai S, 2017. Indian Plate and Its Epic Voyage from Gondwana to Asia: Its Tectonic, Paleoclimatic, and Paleobiogeographic Evolution. *Geological Society of America*, Special Paper 529.

Chen A, White ND, Benson RBJ, Braun MJ and Field DJ, 2019. Total-evidence framework reveals complex morphological evolution in nightbirds (Strisores). *Diversity*, 11, 143.

Chow MC and Qi T, 1978. Paleocenemammalian fossils from Nomogen formation of Inner Mongolia. *Vertebrata PalAsiatica*, 16, 77–85.

Churakov G, Sadasvuni MK, Rodenbloom KR, et al., 2010. Rodent evolution: back to the root. *Molecular Biology and Evolution*, 27, 1315–1326.

Clyde WC and Gingerich PD, 1998. Mammalian community response to the latest Paleocene thermal maximum: An isotaphonomic study in the northern Bighorn Basin, Wyoming. *Geology*, 26, 1011–1014.

Congreve CR and Lamsdell JC, 2016. Implied weighting and its utility in palaeontological datasets: a study using modelled phylogenetic matrices. *Palaeontology*, *59*(3), 447–462.

Cope ED, 1973. *Synopsis of new Vertebrata from the tertiary of Colorado: obtained during the summer of 1873.* US Government Printing Office: USA

Coster P, Benammi M, Lazzari V, Billet G, Martin T, Salem M, Bilal AA, Chaimanee Y, Schuster M, Valentin X, and Brunet M, et al., 2010. *Gaudeamus lavocati* sp. nov. (Rodentia, Hystricognathi) from the early Oligocene of Zallah, Libya: first African caviomorph. *Naturwissenschaften,* 97, 697–706.

Cox PG and Faulkes CG, 2014. Digital dissection of the masticatory muscles of the naked mole-rat, *Heterocephalus glaber* (Mammalia, Rodentia). *PeerJ*, e448.

Cox PG and Jeffery N, 2011. Reviewing the morphology of the jaw-closing musculature in squirrels, rats, and guinea pigs with contrast-enhanced MicroCt. *The Anatomical Record*, 294, 915–928.

Cox PG, Rayfield EJ, Fagan MJ, Herrel A, Pataky TC and Jeffery N, 2012. Functional evolution of the feeding system in rodents. *PLoS ONE*, 7(4), e36299.

D'Ambrosia AR, Clyde WC, Fricke HC, Gingerich PD, Abels HA, 2017. Repetitive mammalian dwarfing during ancient greenhouse warming events. *Science Advances*, 3(3), e1601430.

Dashzeveg D, 1990. New trends in adaptive radiation of early Tertiary rodents (Rodentia, Mammalia). *Acta Zoologica Cracoviensia*, 33, 37–44.

Dashzeveg D and Russell DE, 1988. Paleocene and Eocene Mixodontia (Mammalia, Glires) of Mongolia and China. *Palaeontology,* 31, 129–164.

Dashzeveg DE, Hartenberger JL, Martin T and Legendre S, 1998. A peculiar minute Glires (Mammalia) from the early Eocene of Mongolia. *Bulletin of Carnegie Museum of Natural History*, 34, 194–209.

Dávalos LM, Velazco PM, Warsi OM, Smits PD and Simmons NB, 2014. Integrating incomplete fossils by isolating conflicting signal in saturated and non-independent morphological characters. *Systematic Biology,* 63, 582–600.

Dawson MR, 1968. Middle Eocene rodents (Mammalia) from northeastern Utah. *Annals of Carnegie Museum*, 39,327–370.

Dawson MR, 1977. Late Eocene rodent radiations: North America, Europe and Asia. *Géobios,* 10, 195–209.

Dawson MR, 2015. Emerging perspectives on some Paleogene sciurognath rodents in Laurasia, Chapter 3. In: PG Cox and L Hautier (eds), *Evolution of the rodents: advances in phylogeny, functional morphology and development*. Cambridge University Press: Cambridge.

Dawson MR and Beard CK, 1996. New late Paleocene rodents (Mammalia) from Big Multi Quarry, Washakie Basin, Wyoming. *Palaeovertebrata,* 25, 301–321.

Dawson MR, Li CK and Qi T, 1984. Eocene ctenodactyloid rodents (Mammalia) of eastern and central Asia. *Carnegie Museum of Natural History Special Publication,* 9, 138–150.

Dawson MR, Krishtalka L and Stuck RK, 1990. Revision of the Wind River faunas, early Eocene of central Wyoming. Part 9. The oldest known hystricomorphous rodent (Mammalia: Rodentia). *Annals of Carnegie Museum*, 59, 135–147.

Day WHE, 1986. Analysis of quartet dissimilarity measures between undirected phylogenetic trees. *Systematic Biology*, 35, 325–333.

DeBry R, 2003. Identifying conflicting signal in a multigene analysis reveals a highly resolved tree: the phylogeny of Rodentia (Mammalia). *Systematic Biology*, 52(5), 604–617.

D'Elía G, Fabre P H, and Lessa E P, 2019. Rodent systematics in an age of discovery: recent advances and prospects. *Journal of Mammalogy*, 100(3), 852–871.

Deng T, Wang X, Fortelius M, Li Q, Wang Y, Tseng ZJ, Takeuchi GT, Saylor JE, Säilä LK and Xie G, 2011. Out of Tibet: Pliocene woolly rhino suggests high-plateau origin of Ice Age megaherbivores. *Science*, 333, 1285–1288.

D'Erchia AM, Gissi C, Pesole G, Saccone C and Arnason U, 1996. The guinea-pig is not a rodent. *Nature*, 381, 597–600.

Donoghue MJ, Doyle JA, Gauthier J, Kluge AG and Rowe T, 1989. The importance of fossils in phylogeny reconstruction. *Annual Review of Ecology and Systematics*, 20, 431–460.

dos Reis M, Inoue J, Hasegawa M, Asher RJ, Donoghue MJ et al., 2012. Phylogenomic datasets provide both precision and accuracy in estimating the timescale of placental mammal phylogeny. *Proceedings of the Royal Society B*, 279, 3491–3500.

Douzery EJP, Delsuc F, Stanhope MJ and Huchon D, 2003. Local molecular clocks in three nuclear genes: divergence times for rodents and other mammals and incompatibility among fossil calibrations. *Journal of Molecular Evolution*, 57, S201–S213.

Doyle JA and Donoghue MJ, 1993. Phylogenies and Angiosperm diversification. *Paleobiology*, 19, 141–167.

Dupont-Nivet G, Krijgsman W, Langereis CG, Abels HA, Dai S and Fang X, 2007. Tibetan Plateau aridification linked to global cooling at the Eocene-Oligocene transition. *Nature,* 445, 635–638.

Edgecombe, GD, 2010. Palaeomorphology: fossils and the inference of cladistics relationships. *Acta Zoologica*, 91, 72–80.

Emry RJ, 1975. Revised Tertiary stratigraphy and paleontology of the western Beaver Divide, Fremont County, Wyoming. *Smithsonian Contributions to Paleobiology*, 25, 1–20.

Emry RJ, 1981. Additions to the mammalian fauna of the type Duchesnean, with comments on the status of the Duchesnean "age". *Journal of Paleontology*, 55(3), 563–570.

Emry RJ, 2007. The middle Eocene North American myomorph rodent *Elymys*, her Asian sister *Aksyiromys*, and other Eocene myomorphs. *Bulletin of Carnegie Museum of Natural History*, 39, 141–150.

Emry RJ and Dawson MR, 1972. A unique cricetid (Rodentia, Mammalia) from the early Oligocene of Natrona County, Wyoming. *American Museum Novitates*, 2508, 1–14.

Emry RJ and Korth WW, 1989. Rodents of the Bridgerian (middle Eocene) Elderberry Canyon local fauna of eastern Nevada. *Smithsonian Contributions to Paleobiology*, 67, 1–14.

Esselstyn JA, Oliveros CH, Swanson MT and Faircloth BC, 2017. Investigating difficult nodes in the placental mammal tree with expanded taxon sampling and thousands of ultraconserved elements, *Genome Biology and Evolution*, 9(9), 2308–2321.

Estabrook GF, McMorris FR and Meacham CA, 1985. Comparison of undirected phylogenetic trees based on subtrees of four evolutionary units. *Systematic Zoology,* 34, 193–200.

Evans AR, Wilson GP, Fortelius M and Jernvall J, 2007. High-level similarity of dentitions in carnivorans and rodents. *Nature,* 445, 78–81.

Fabre PH, Hautier L, Dimitrov D and Douzery EJ, 2012. A glimpse on the pattern of rodent diversification: a phylogenetic approach. *BMC Evolutionary Biology*, 12(88), 1–9.

Fabre PH, Hautier L and Douzery EJ, 2015. A synopsis of rodent molecular phylogenetics, systematics and biogeography, 19-69. In: PG Cox and L Hautier (eds), *Evolution of the rodents: advances in phylogeny, functional morphology and development*. Cambridge University Press: Cambridge.

Fahlbusch V, 1985. Origin and evolutionary relationships among geomyoids, 617-629. In Luckett WP and Hartenberger J (eds), *Evolutionary relationships among rodents: a multidisciplinary analysis.* Springer: Boston.

Faircloth BC, McCormack JE, Crawford NG, Harvey MG, Brumfield RT and Travis CG, 2012. Ultraconserved elements anchor thousands of genetic markers spanning multiple evolutionary timescales. *Systematic Biology,* 61(5), 717–726.

Farris JS, 1969. A successive approximations approach to character weighting. *Systematic Biology*, *18*(4), 374–385.

Flynn, JJ and Wyss, 1998. Recent advances in South American mammalian paleontology. *Trends in Ecology and Evolution*, 13, 449–454.

Flynn JJ, Wyss AR, Croft DA and Charrier R, 2003. The Tinguiririca fauna, Chile: biochronology, paleoecology, biogeography, and a new earliest Oligocene South American Land Mammal 'Age'. *Palaeogeography, Palaeoclimatology, Palaeoecology*, 194(3–4), 229–259.

Flynn LJ, 2008. Hystricognathi and Rodentia incertae sedis, 498–506. In: Janis CM, Gregg GF and Uhen MD (eds), *Evolution of tertiary mammals in North America. Small mammals, xenarthrans, and marine mammals (vol 2).* Cambridge University Press: Cambridge.

Foley NM, Springer MS and Teeling EC, 2016. Mammal madness: is the mammal tree of life not yet resolved? *Philosophical Transactions of the Royal Society B: Biological Sciences*, 371(1699), 20150140.

Fostowicz-Frelik Ł, 2020. Most Successful Mammals in the Making: A Review of the Paleocene Glires, chapter 5. In: Pontarotti P (ed), *Evolutionary Biology—A Transdisciplinary Approach*. Springer, Cham.

Fostowicz-Frelik Ł, López-Torres S and Li Q, 2021. Tarsal morphology of ischyromyid rodents from the middle Eocene of China gives an insight into the group's diversity in Central Asia. *Scientific Reports*, 11(1), 11543.

Gatesy J and Arctander P, 2000. Hidden morphological support for the phylogenetic placement of *Pseudoryx nghetinhensis* with bovine bovids: a combined analysis of gross anatomical evidence and DNA sequences from five genes. *Systematic Biology*, 49, 515–538.

Gatesy J and Baker RH, 2005. Hidden Likelihood support in genomic data: can forty-five wrongs make a right? *Systematic Biology,* 54(3), 483–492.

Gatesy J, O'Grady P and Baker RH, 1999. Corroboration among data sets in simultaneous analysis: hidden support for phylogenetic relationships among higher level artiodactyl taxa. *Cladistics*, 15, 271–313.

Giles S, Xu G H, Near T J and Friedman M, 2017. Early members of 'living fossil' lineage imply later origin of modern ray-finned fishes. *Nature*, 549(7671), 265–268.

Goloboff PA, 1993. Estimating characters weights during tree search. *Cladistics,* 9, 83–91.

Goloboff PA, 1997. Self-weighted optimization: tree searches and character state reconstructions under implied transformation costs. *Cladistics*, 13, 225–245.

Goloboff PA and Catalano SA, 2016. TNT version 1.5, including a full implementation of phylogenetic morphometrics. *Cladistics,* 32, 221–238.

Goloboff PA, Carpenter JM, Arias JS and Esquivel DRM, 2008. Weighting against homoplasy improves phylogenetic analysis of morphological datasets. *Cladistics*, 24, 758–773.

Goloboff PA, Mirande JM and Salvador JS, 2010. On weighting characters differently in different parts of the cladogram. *Cladistics* 26, 212.

Goloboff PA, Pittman M and Xu X, 2019. Morphological data sets fit a common mechanism much more poorly than DNA sequences and call into question the Mkv model. *Systematic Biology*, 68(3), 494–504.

Goswami A, Milne N and Wroe S, 2011. Biting through constraints: cranial morphology, disparity and convergence across living and fossil carnivorous animals. *Proceedings of the Royal Society London B, Biological Sciences,* 278, 1831–1839.

Hahn DH, and Manabe S, 1975. The role of mountains in the south Asian monsoon circulation. *Journal of the Atmospheric Sciences*, 321, 1515–1541.

Halanych KM, 2016. How our view of animal phylogeny was reshaped by molecular approaches: lessons learned. *Organisms Diversity & Evolution*, 16, 319–328.

Hallstrom BM and Janke A, 2010. Mammalian evolution may not be strictly bifurcating. *Molecular Biology and Evolution,* 27, 2804–2816.

Hartenberger JL, 1985. The Order Rodentia: major questions on their evolutionary origin, relationships and suprafamilial systematics, 1–34. In: Luckett WP and Hartenberger JL (eds), *Evolutionary relationships among rodents: A multidisciplinary analysis.* Plenum Press: New York.

Hautier L, Michaux J, Marivaux L, and Vianey-Liaud M, 2008. Evolution of the zygomasseteric construction in Rodentia, as revealed by a geometric morphometric analysis of the mandible of *Graphiurus* (Rodentia, Gliridae). *Zoological Journal of the Linnean Society,* 154(4), 807–821.

Hautier L, Lebrun R, Saksiri S, Michaux J, Vianey-Liaud M, and Marivaux L, 2011. Hystricognathy vs sciurognathy in the rodent jaw: a new morphometric assessment of hystricognathy applied to the living fossil *Laonastes* (Diatomyidae). *PLoS ONE*, 6(4), e18698.

Hautier L, Cox PG and Lebrun R, 2012. Grades and clades among rodents: the promise of geometric morphometrics, 277–299 In: PG Cox and L Hautier (eds), *Evolution of the rodents: advances in phylogeny, functional morphology and development*. Cambridge University Press: Cambridge.

Heaton TH, 1996. Ischyromyidae, 373–398. In: Prothero DR and Emry RJ (eds), *The Terrestrial Eocene-Oligocene Transition in North America*. Cambridge University Press: Cambridge.

Hecke TV, 2012. Power study of anova versus Kruskal-Wallis test. *Journal of Statistics and Management Systems*, 15(2–3), 241–247.

Hennig W, 1981. *Insect phylogeny*. Wiley: New York.

Houlle A, 1999. The origin of platyrrhines: an evaluation of the Antarctic scenario and the floating island model. *American Journal of Physical Anthropology,* 109, 541–559.

Huang X, Li C, Dawson MR and Liu L, 2004. *Hanomys malcolmi*, a new simplicidentate mammal from the Paleocene of central China: its relationships and stratigraphic implications. *Bulletin of Carnegie Museum of Natural History*, 36, 81–9.

Huchon D and Douzery EJP, 2001. From the Old World to the New World: a molecular chronicle of the phylogeny and biogeography of hystricognath rodents. *Molecular Phylogenetics and Evolution*, 20, 238–251.

Huchon D, François M, Catzeflis M and Douzery EJP, 2000. Variance of molecular datings, evolution of rodents and the phylogenetic affinities between Ctenodactylidae and Hystricognathi. *Proceedings of the Royal Society London B,* 267, 393–402.

Huchon D, Chevret P, Jordan U, Kilpatrick CW, Ranwez V, Jenkins PD, Brosius J and Schmitz J, 2007. Multiple molecular evidences for a living mammalian fossil. *PNAS*, 104(18), 7495–7499.

Ivy LD, 1990. Systematics of the late Paleocene and early Eocene Rodentia (Mammalia) from the Clarks Fork Basin, Wyoming. *Contributions from the Museum of Paleontology University of Michigan*, 28, 21–70.

Ji Q, Luo Z-X and Ji S-A, 1999. A Chinese triconodon mammal and mosaic evolution of the mammalian skeleton. *Nature*, 398, 326–330.

Jiang D, Klaus S, Zhang YP, Hillis DM and Li JT, 2019. Asymmetric biotic interchange across the Bering land bridge between Eurasia and North America. *National Science Review*, 6(4), 739–745.

Johnson M, Zaretskaya I, Raytselis Y, Merezhuk Y, McGinnis S and Madden TL, 2008. NCBI BLAST: a better web interface. *Nucleic Acids Research*, 36, W5–W9.

Katoh K, Misawa K, Kuma K, Miyata T. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Research*, 30(14), 3059–3066.

Kent J, 2014. faSomeRecords: Linux binary available from UCSC. http://hgdownload.cse.ucsc.edu/admin/exe/linux.x86_64/

Kimura Y, Hawkins MTR, McDonough MM, Jacobs LL and Flynn LJ, 2015. Corrected placement of *Mus-Rattus* fossil calibration forces precision in the molecular tree of rodents. *Scientific Reports*, 5, 1–9.

Kitazoe Y, Kishino H, Waddell PJ, Nakajima N, Okabayashi T et al., 2007. Robust time estimation reconciles views of the antiquity of placental mammals. *PLoS ONE*, 2, e384.

Klingenberg CP, 2008. Morphological integration and development modularity. *Annual Review of Ecology*, *Evolution and Systematics*, 39, 115–132.

Klingenberg CP, 2011. MorphoJ: an integrated software package for geometric morphometrics. *Molecular Ecology Resources*, 11, 353–357.

Koch NM and Parry LA, 2020. Death is on our side: paleontological data drastically modify phylogenetic hypotheses. *Systematic biology*, 69(6), 1052–1067.

Korth WW, 1984. Earliest Tertiary evolution and radiation of rodents in North America. *Bulletin of Carnegie Museum of Natural History,* 24, 1–71.

Korth WW, 1994. *The Tertiary record of rodents in North America*. Plenum Press: New York.

Krause DW, 1982. Jaw movement, dental function, and diet in the Paleocene multituberculate *Ptilodus*. *Paleobiology*, 8(3), 265–281.

Kumar S and Hedges SB, 1998. A molecular timescale for vertebrate evolution. *Nature*, 392(6679), 917–920.

Landfear R, Frandsen PB, Wright AM, Senfeld T and Calcott B, 2017. PartitionFinder2: new methods for selecting partitioned models of evolution for molecular and morphological and morphological phylogenetic analysis. Molecular *Biology and Evolution*, 34(3), 772–773.

Landis MJ, Matzke NJ, Moore BR and Huelsenbeck JP, 2013. Bayesian analysis of biogeography when the number of areas is large. *Systematic Biology*, 62, 789–804.

Landry SO Jr, 1957. The Interrelationships of the New and Old World Hystricomorph Rodents. *University of California Publications in Zoology*, 56, 1–118.

Landry, SO Jr, 1999. A proposal for a new classification and nomenclature for the Glires (Lagomorpha and Rodentia). *Zoosystematics and Evolution*, 75, 283–316.

Lewis PO, 2001. A likelihood approach to estimating phylogeny from discrete morphological character data. *Systematic Biology*, 50(6), 913–925.

Li CK, 1977. Paleocene eurymylids (Anagalida, Mammalia) of Qianshan, Anhui. *Vertebrata PalAsiatica*, 15(2), 103–118.

Li CK and Ting SY, 1985. Possible phylogenetic relationship of Asiatic eurymylids and rodents, with comments on mimotonids, 35–58. In: WP Luckett and JL Hartenberger (eds), *Evolutionary relationships among rodents*. Springer: Boston.

Li CK and Ting SY, 1993 New cranial and postcranial evidence for the affinities of the eurymylids (Rodentia) and mimotonids (Lagomorpha), 151–158. In: FS Szalay, MJ Novacek, MC McKenna (eds), *Mammal phylogeny—placentals*. Springer: Berlin.

Li CK, Zheng JJ and Ting SY, 1989. The skull of *Cocomys lingchaensis*, an early Eocene ctenodactyloid rodent of Asia. *Natural History Museum of Los Angeles County Science Series,* 33, 179–192.

Li CK, Wang YQ, Zhang ZQ, Mao FY and Meng J, 2016. A new mimotonidan mammal *Mina hui* (Mammalia, Glires) from the Middle Paleocene of Qianshan, Anhui Province, China. *Vertebrata Palasiatica,* 54, 121–136.

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durnbin R and 1000 Genome Project Data Processing Subgroup, 2009. The Sequence alignment/map (SAM) format and SAMtools. *Bioinformatics*, 25, 2078–9.

Li Q, 2016. Eocene fossil rodent assemblages from the Erlian Basin (Inner Mongolia, China): biochronological implications. *Palaeoworld*, 25(1), 95-103.

Li Q, 2020. New late Eocene cylindrodontid rodents from the Erlian Basin (Nei Mongol, China). *Palaeobiodiversity and Palaeoenvironments,* 100(4), 1083-1094.

Li Q, 2021. Additional tsaganomyid, cylindrodontid and ctenodactyloid rodent materials from the Erden Obo section, Erlian Basin (Nei Mongol, China). *Vertebrata PalAsiatica*, 59(1), 1-18.

Li Q and Wang X, 2015. Into Tibet: An Early Pliocene Dispersal of Fossil Zokor (Rodentia: Spalacidae) from Mongolian Plateau to the Hinterland of Tibetan Plateau. *PLoS ONE*, 10(2), e0144993.

Li Y, Zhang Y, Wu X, Ao H, Li L and A Z, 2014. Mammalian evolution in Asia linked to climate changes, 435–490. In An Z (ed), *Late Cenozoic climate change in Asia*. Springer: Dordrecht.

Limaye A, 2012. Drishti – volume exploration and presentation tool. *Proceedings SPIE Developments in X-ray Tomography VIII,* 85060X.

Lin C, Wu H, Ou T, Chen D, 2019. A new perspective on solar dimming over the Tibetan Plateau. *International Journal of Climatology*, 39(1), 302-16.

Lin Y, Rajan V and Moret BM, 2012. A metric for phylogenetic trees based on matching. *IEE/ACM Transactions on Computational Biology*, 9(4), 1014–1022.

Liu H, Sun Z, Wang J and Min J, 2004. A modeling study of the effects of anomalous snow cover over the Tibetan Plateau upon the South Asian summer monsoon. *Advances in Atmospheric Sciences*, 21, 964–975.

Lopatin AV and Averianov AO, 2004a. The earliest rodents of the genus *Tribosphenomys* from the Paleocene of Central Asia. *Doklady Biological Sciences*, 397, 336–337.

Lopatin AV and Averianov AO, 2004b. A new species of *Tribosphenomys* (Mammalia: Rodentiaformes) from the Paleocene of Mongolia. *New Mexico Museum of Natural History and Science Bulletin*, 26, 169–175.

Lopatin A and Kondrashov P, 2003. The skull structure of *Sinomylus* (Mixodontia). *Journal of Vertebrate Paleontology,* 23(3), 72A–73A.

Luckett WP and Hartenberger J-L, 1985. *Evolutionary Relationships among Rodents: A Multidisciplinary Analysis*. Springer: New York.

Lukashevich ED and Ribeiro GC, 2019. Mesozoic fossils and the phylogeny of Tipulomorpha (Insecta: Diptera). *Journal of Systematic Palaeontology*, *17*(8), 635–652.

Maddison WP and Maddison DR, 2019. Mesquite: a modular system for evolutionary analysis. Version 3.61 http://www.mesquiteproject.org

Magallón SA, 2004. Dating lineages: molecular and paleontological approaches to the temporal framework of clades. *International Journal of Plant Sciences*, 164, S7–S21.

Magallón SA and Sanderson MJ, 2001. Absolute diversification rates in angiosperm clades. *Evolution*, 55, 1762–1780.

Maier W and Schrenk F, 1987. The hystricomorphy of the Bathyergidae, as determined from ontogenetic evidence. *Zeitschrift für Säugetierkunde*, 52, 156–164.

Maier W, Klingler P and Ruf I, 2002. Ontogeny of the medial masseter muscle, pseudo-myomorphy, and the systematic position of the Gliridae (Rodentia, Mammalia). *Journal of Mammalian Evolution*, 9, 253–269.

Mao FY, Li Q, Wang YQ and Li CK, 2017. *Taizimylus tongi*, a new eurymylid (Mammalia, Glires) from the upper Paleocene of Xinjiang, China. *Palaeoworld*, 26, 519–530.

Marivaux L and Boivin M, 2019. Emergence of hystricognathous rodents: Palaeogene fossil record, phylogeny, dental evolution and historical biogeography. *Zoological Journal of the Linnean Society*, 187(3), 929–964.

Marivaux L, Vianey-Liaud M and Jaeger J-J, 2004. High-level phylogeny of Tertiary rodents: dental evidence. *Zoological Journal of the Linnean Society,* 142, 105–134.

Marivaux L, Essid EM, Marzougui W, Khayati Ammar H, Adnet S, Marandat B, Merzeraud G, Tabuce R and Vianey-Liaud M, 2014. A new and primitive species of *Protophiomys* (Rodentia, Hystricognathi) from the Late Middle Eocene of Djebel el Kébar, Central Tunisia. *Palaeovertebrata*, 38, 1–17.

Marivaux L, Adnet S, Benammi M, Tabuce R and Benammi M, 2016. Anomaluroid rodents from the earliest Oligocene of Dakhla, Morocco, reveal the long-lived and morphologically conservative pattern of the Anomaluridae and Nonanomaluridae during the Tertiary in Africa. *Journal of Systematic Palaeontology,* 15(7), 539–569.

Martin AR and Richardson G, 2017. Rodent eradication scaled up: clearing rats and mice from South Georgia. *Oryx*, 53(1), 27–35.

Mason VC, Li G, Minx P, Schmitz J, Churakov G, Doronina L, Melin AD, Dominy NJ, Lim NT, Springer MS and Wilson RK, 2016. Genomic analysis reveals hidden biodiversity within colugos, the sister group to primates. *Science Advances*, 2, e1600633.

Matthew WD and Granger W, 1925. Fauna and correlation of the Gashato Formation of Mongolia. *American Museum novitates*, 189, 1–12.

Matzke NJ, 2013. Probabilistic historical biogeography: new models for founder-event speciation, imperfect detection, and fossils allow improved accuracy and model-testing. *Frontiers of Biogeography,* 5(4), 242–248.

McCormack JE, Faircloth BC, Crawford NG, Gowaty PA, Brumfield RT and Glenn TC, 2012. Ultraconserved elements are novel phylogenomic markers that resolve placental mammal phylogeny when combined with species-tree analysis. *Genome Research,* 22, 746–754.

McInerney FA and Wing SL, 2011. The Paleocene-Eocene Thermal Maximum: A perturbation of carbon cycle, climate, and biosphere with implications for the future. *Annual Review of Earth and Planetary Sciences,* 39, 489–516.

McKenna MC and Meng J, 2001. A primitive relative of rodents from the Chinese Paleocene. *Journal of Vertebrate Paleontology*, 21, 565–572.

McLean BS, Bell KC, Allen JM, Helgen KM and Cook JA, 2019. Impacts of inference method and data set filtering on phylogenomic resolution in a rapid radiation of ground squirrels (Xerinae: Marmotini). *Systematic Biology*, 68, 298–316.

Meng J and Wyss AR, 2001. The morphology of Tribosphenomys (Rodentiaformes, Mammalia): phylogenetic implications for basal Glires. *Journal of Mammalian Evolution,* 8, 1–71.

Meng J, Wyss AR, Dawson MR and Zhai R, 1994. Primitive fossil rodent from Inner Mongolia and its implications for mammalian phylogeny. *Nature*, 370, 134–136.

Meng J, Zhai RJ and Wyss AR, 1998. The late Paleocene Bayan Ulan fauna of Inner Mongolia, China. *Bulletin of Carnegie Museum of Natural History*, 34, 148–85.

Meng J, Hu Y and Li C, 2003. The osteology of Rhombomylus (Mammalia, Glires): implications for phylogeny and evolution of Glires. *Bulletin of the American Museum of Natural History,* 1–247.

Meng J, Wyss AR, Hu Y, Wang Y, Bowen GJ and Koch PL, 2005. Glires (Mammalia) from the Late Paleocene Bayan Ulan locality of Inner Mongolia. *American Museum Novitates*, 3473, 1–25.

Meng J, Ni X, Li C, Beard K, Gebo DL, Wang Y and Wang H, 2007. New material of Alagomyidae (Mammalia, Glires) from the Late Paleocene Subeng locality, Inner Mongolia. *American Museum Novitates,* 3597, 1–29.

Meng J, Wang Y, Ni X, Beard KC, Sun C, Li Q, Jin X and Bai B, 2007. New stratigraphic data from the Erlian Basin: implications for the division, correlation, and definition of Paleogene lithological units in Nei Mongol (Inner Mongolia). *American Museum Novitates*, 3570, 1–31.

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

Millien V, 2008. The largest among the smallest: the body mass of the giant rodent *Josephoartigasia monesi*. *Proceedings of the Royal Society B: Biological Sciences*, 275(1646), 1953–1955.

Montgelard C, Forty E, Arnal V and Matthee CA, 2008. Suprafamilial relationships among Rodentia and the phylogenetic effect of removing fast-evolving nucleotides in mitochondrial, exon and intron fragments. *BMC Evolutionary Biology,* 8(321), 1–16.

Morris PJR, Cobb SNF and Cox PG, 2018. Convergent evolution in the Euarchontoglires. *Biology Letters,* 14(8), 20180366.

Mounce RCP, Sansom R and Wills MA, 2016. Sampling diverse character improves phylogenies: craniodental and postcranial characters of vertebrates often imply different trees. *Evolution*, 70, 666–686.

Oliver O and Daxner-Höck G, 2017. Large-sized species of Ctenodactylidae from the Valley of Lakes (Mongolia): an update on dental morphology, biostratigraphy, and paleobiogeography. *Palaeontologia Electronica*, 20(1), 20.1.1A.

Oliver A, Sanisidro O, Baatarjav B, Niiden I and Daxner-Höck, 2019. Diversification rates in Ctenodactylidae (Rodentia, Mammalia) from Mongolia. *Palaeobiodiversity and Palaeoenvironments,* 97, 51–65.

O'Reilly JE, Puttick MN, Parry L, Tanner AR, Tarver JE, Fleming J, Pisani D and Donoghue PCJ, 2016. Bayesian methods outperform parsimony but at the expense of precision in the estimation of phylogeny from discrete morphological data. *Biology Letters*, 12, 20160081.

Owen R, 1870. XXIII. On the fossil mammals of Australia– Part III. *Diprotodon australis*, Owen. *Philosophical Transactions of the Royal Society of London*, 160, 519–578.

Patterson C, 1981. Significance of fossils in determining evolutionary relationships. *Annual Review of Ecology and Systematics*, 12, 195–223.

Pattinson DJ, Thompson RS, Piotroeski AK and Asher RJ, 2014. Phylogeny, paleontology, and primates: do incomplete fossils bias the tree of life? *Systematic Biology*, 64(2), 169–186.

Pisano J, Condamine FL, Lebedev V, Bannikova A, Quéré J, Shenbrot GI, Pagès M and Michaux JR, 2015. Out of Himalaya: the impact of past Asian environmental changes on the evolutionary and biogeographical history of Dipodoidea (Rodentia). *Journal of Biogeography*, 42, 856–870.

Polly PD, Cardini A, Davis EB and Steppan SJ, 2015. Marmot evolution and global change in the past 10 million years, 246–276. In: Cox PG and Hautier L (eds), *Evolution of the Rodents: Advances in Phylogeny, Palaeontology and Functional Morphology*. Cambridge University Press: Cambridge.

Portik DM and Wiens JJ, 2021. Do alignment and trimming methods matter for phylogenomic (UCE) analyses? *Systematic biology*, 70(3), 440–62.

Poux C, Chevret P, Huchon D, de Jong WW and Douzery EJ, 2006. Arrival and diversification of caviomorph rodentsa and platyrrhine primates in South America. *Systematic Biology*, 55(2), 228–244.

Prebus MM, 2021. Phylogenomic species delimitation in the ants of the *Temnothorax salvini* group (Hymenoptera: Formicidae): an integrative approach. *Systematic Entomology*, 46(2), 307–26.

Prothero DR, 1996. Magnetic Stratigraphy of the White River Group in the High Plains, 262. In: Prothero DR and Emry RJ (eds), *The Terrestrial Eocene-Oligocene transition in North America*. Cambridge University Press: Cambridge.

Prothero DR and Emry RJ, 1996. *The Terrestrial Eocene-Oligocene Transition in North America.* Cambridge University Press: Cambridge.

Prothero DR and Emry RJ, 2004. The Chadronian, Orellan, and Whitneyan North American Land Mammal Ages, 156–168. In: Woodburne MO (ed.), *Late Cretaceous and Cenozoic Mammals of North America*. Columbia University Press: New York.

Puttick MN, O'Reilly JE, Pisani D and Donoghue PCJ, 2018. Probabilistic methods outperform parsimony in the phylogenetic analysis of data simulated without a probabilistic model. *Palaeontology*, 62(1), 1–17.

Qiu L, 2008. The Third Pole. *Nature*, 454, 393–396.

Qiu Z and Li C, 2005. Evolution of Chinese mammalian faunal regions and elevation of the Qinghai-Xizang (Tibet) Plateau. *Science in China Series D: Earth Sciences*, *48*(8), 1246–1258.

R Core Team, 2020. R: a language and environment for statistical computing. *R Foundation for Statistical Computing*, https://www.R-project.org.

Rambaut A, Drummond AJ, Xie D, Baele G and Suchard MA, 2018. Posterior summarisation in Bayesian phylogenetics using Tracer 1.7. *Systematic Biology*, 67(5), 901–904.

Rankin AH, Emry RJ and Asher RJ, 2020. Anatomical sciuromorphy in "protrogomorph" rodents. *Palaeontologia Electronica,* 23, 1–26.

Raup DM, 1972. Taxonomic diversity during the Phanerozoic. *Science*, 177, 1065–1071.

Rinderknecht A and Blanco RE, 2008. The largest fossil rodent. *Proceedings of the Royal Society B: Biological Sciences*, 275(1637), 923–928.

Robinson DF and Foulds LR, 1981. Comparison of phylogenetic trees. *Mathematical Biosciences*, 53(1–2), 131–147.

Rodrigues HG, Marivaux L and Vianey-Liaud M, 2010. Phylogeny and systematic revision of Eocene Cricetidae (Rodentia, Mammalia) from Central and East Asia: on the origin of cricetid rodents. *Journal of Evolutionary Zoological and Evolutionary Research*, 48 (3), 259–268.

Rodríguez-Prieto A, Igea J and Castresana J, 2014. Development of rapidly evolving intron markers to estimate multilocus species trees of rodents. *PLoS ONE*, 9, e96032.

Ronquist F, 1997. Dispersal-vicariance analysis: A new approach to the quantification of historical biogeography. *Systematic Biology*, 46, 195–203.

Ronquist F and Huelsenbeck JP, 2003. MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics*, 19(12), 1572–1574.

Ronquist F, Klopfstein S, Vilhelmson L, Schulmeister S, Murray DL and Rasnitsyn AP, 2012. A total-evidence approach to dating with fossils, applied to the early radiation of the Hymenoptera. *Systematic Biology*, 61, 973–999.

Rose KD, 1981. The Clarkforkian land-mammal age and mammalian composition across the Paleocene-Eocene boundary. *University of Michigan Papers on Paleontology*, 26, 1–197.

Royden LH, Burchfiel CB and van der Hilst RD, 2008. The Geological Evolution of the Tibetan Plateau. *Science,* 321(5892), 1054–1058.

Rüber L, Tan HH, Britz R. Snakehead (Teleostei: Channidae) diversity and the Eastern Himalaya biodiversity hotspot, 2020. *Journal of Zoological Systematics and Evolutionary Research*, 58(1), 356–86.

Sansom RS and Wills AW, 2013. Fossilization causes organisms to appear erroneously primitive by distorting evolutionary trees. *Scientific Reports*, 66(5), 813–822.

Sansom RS, Gabbott SE and Purnell AM, 2010. Non-random decay of chordate characters causes bias in fossil interpretation. *Nature*, 463, 797–800.

Sansom RS, Wills MA and Williams T, 2017. Dental data perform relatively poorly in reconstructing mammal phylogenies: morphological partitions evaluated with molecular benchmarks. *Systematic Biology,* 66(5), 813–822.

Sansom RS, Choate PG, Keating JN and Randle E, 2018. Parsimony, not Bayesian analysis, recovers more stratigraphically congruent phylogenetic trees. *Biology Letters*, 14, 20180263.

Sato JJ, Bradford TM, Armstrong KN, Donnellan SC, Echenique-Diaz LM, Begué-Quiala G, Gámez-Díez J, Yamaguchi N, Nguyen ST, Kita M and Ohdachi SD, 2019. Post K-Pg diversification of the mammalian

order Eulipotyphla as suggested by phylogenomic analyses of ultra-conserved elements. *Molecular Phylogenetics and Evolution,* 141, 106605.

Scarpetta G, 2020. Combined-evidence analyses of ultraconserved elements and morphological data: an empirical example in iguanian lizards. *Biology Letters,* 16, 20200356.

Scotese CR, 2016. PALEOMAP PaleoAtlas for GPlates and the PaleoData Plotter Program. PALEOMAP Project, http://www.earthbyte.org/paleomap-paleoatlas-for-gplates/

Scotland RW, Olmstead RG and Bennett JR, 2003. Phylogeny reconstruction: the role of morphology. *Systematic Biology*, 52(4), 539–548.

Secord R, Bloch JI, Chester SGB, Boyer DM, Wood AR, Wing SL, Kraus MJ, McInerney FA and Krigbaum J, 2012. Evolution of the earliest horses driven by climate change in the Paleocene-Eocene thermal maximum. *Science*, 335, 959–962.

Shevyreva NS, 1984. New early Eocene rodents from the Zaysan Basin, 114. In: Gabunia LK (ed), *Flora I fauna Zaysanskoi vpadiny*. Akademiya Nauk Gruzinskoy: Russia.

Shih HT, Yeo DC, Ng PK. The collision of the Indian plate with Asia: molecular evidence for its impact on the phylogeny of freshwater crabs (Brachyura: Potamidae), 2009. *Journal of Biogeography*, 36(4), 703–19.

Shu D-G, Luo H-L, Conway Morris S, Zhang X-L, Hu S-X, Chen L, Han, J, Zhu M, Li Y and Chen L-Z, 1999. Lower Cambrian vertebrates from south China. *Nature*, 402, 42–46.

Silvestro M, 2012. raxmlGUI: a graphical front-end for RAxML. *Organisms Diversity and Evolution*, 12, 335–337.

Simmons MP, 2012. Misleading results of likelihood-based phylogenetic analyses in the presence of missing data. *Cladistics*, 28(2), 208–222.

Simmons MP, 2014. A confounding effect of missing data on character conflict in maximum likelihood and Bayesian MCMC phylogenetic analyses. *Molecular phylogenetics and evolution*, 80, 267–280.

Simpson GG, 1945. The principles of classification and a classification of mammals. *Bulletin of the American Museum of Natural History*, 85, 1–350.

Simpson GG, 1952. How many species? *Evolution*, 6, 342–342.

Sloan RE, 1969. Cretaceous and Paleocene terrestrial communities of western North America. *Proceedings of the North American Paleontological Convention (Part E*), 1, 427–453.

Smith AB, 1998. What does palaeontology contribute to systematics in a molecular world? *Molecular Phylogenetics and Evolution*, 9(3), 437–447.

Smith MR, 2019. Bayesian and parsimony approaches reconstruct informative trees from simulated morphological datasets. *Biology Letters*, 15, 20180632.

Smith SA and O'Meara BC, 2012. treePL: divergence time estimation using penalized likelihood for large phylogenies. *Bioinformatics*, 28(20), 2689–2690.

Springer MS, Burk-Herrick A, Meredith R, Eizirik E, Teeling E, O'Brien SJ and Murphy WJ, 2007. The adequacy of morphology for reconstructing the early history of placental mammals. *Systematic Biology,* 56, 673–684.

Stamatakis A, 2014. RAxML Version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics,* 30(9), 1312–1313.

Storch G and Seiffert C, 2007. Extraordinarily preserved specimen of the oldest known glirid from the middle Eocene of Messel (Rodentia). *Journal of Vertebrate Paleontology*, 27(1), 189–194.

Sun, X and Wang P, 2005. How old is the Asian monsoon system? – palaeobotanical records from China. *Palaeogeography, Palaeoclimatology and Palaeoecology*, 222(3–4), 118–222.

Swanson MT, Oliveros CH and Esselstyn JA, 2019. A phylogenomic rodent tree reveals the repeated evolution of masseter architectures. *Proceedings of the Royal Society B,* 286, 20190672.

Swofford DL, 2003. PAUP*. Phylogenetic Analysis Using Parsimony (*and Other Methods). Version 4. Sinauer Associates, Sunderland, Massachusetts.

Sych L, 1971. Mixodontia, a new order of mammals from the Paleocene of Mongolia. *Palaeontologica Polonica*, 25, 147–158.

Tabrum AR, Prothero DR, Garcia D and Emry RJ, 1996. Magnetostratigraphy and biostratigraphy of the Eocene-Oligocene transition, southwestern Montana, 278–311. In: Prothero DR and Emry RJ (eds), *The Terrestrial Eocene-Oligocene transition in North America*. Cambridge University Press: Cambridge.

Tarver JE, Dos Reis M, Mirarab S, Moran RJ, Parker S, O'Reilly JE, King BL, O'Connell MJ, Asher RJ, Warnow T and Peterson KJ, 2016. The interrelationships of placentals and the limits of phylogenetic inference. *Genome Biology and Evolution*, 8, 330–344.

Thompson RS, Bärmann EV and Asher RJ, 2012. The interpretation of hidden support in combined data phylogenetics. *Journal of Zoological Systematics and Evolutionary Research*, *50*(4), 251–263.

Tong Y, 1977. *Pappocricetodon*, a pre-Oligocene cricetid genus (Rodentia) from central China. *Vertebrata PalAsiatica*, 30, 1–16.

Tong Y and Dawson MR, 1995. Early Eocene rodents (Mammalia) from Shandong Province, People's Republic of China. *Annals of Carnegie Museum*, 64, 51–63.

Torsvik TH and Cocks LRM, 2017. *Earth History and Palaeogeography*. Cambridge University Press: Cambridge.

Tseng ZJ, Li Q and Wang X, 2013a. A new cursorial hyena from Tibet, and analysis of biostratigraphy, paleozoogeography, and dental morphology of Chasmaporthetes (Mammalia, Carnivora). *Journal of Vertebrate Paleontology*, 33, 1457–1471.

Tseng ZJ, Wang X, Slater GJ, Takeuchi GT, Li Q, Liu J and Xie G, 2013b. Himalayan fossils of the oldest known pantherine establish ancient origin of big cats. *Proceedings of the Royal Society B: Bological Sciences*, 281, 20132686.

Tucker DB, Hedges SB, Colli GR, Pyron RA and Sites Jr JW, 2017. Genomic timetree and historical biogeography of Carribean island Ameiva lizards (*Pholidoscelis*: Teiidae). *Ecology and Evolution*, 7, 7080–7090.

Tukey JW, 1949. Comparing individual means in the analysis of variance. *Biometrics*, 5(2), 99–114.

Tullberg T, 1899. Ueber das system der nagethiere : ein phylogenetische studie. *Nova Acta Regiae Societatis Scientarium Upsaliersis*, 18, 1–514.

Veizer J, Ala D, Azmy K, Bruckschen P, Buhl D, Bruhn F, Carden GA, Diener A, Ebneth S, Godderis Y and Jasper T, 1999. $^{87}Sr/^{86}Sr$, $\delta^{13}C$ and $\delta^{18}O$ evolution of Phanerozoic seawater. *Chemical Geology,* 161, 59–88.

Veniaminova NA, Vassetzky NS, Lavrenchenko LA, Popov SV and Kramerov DA, 2007. Phylogeny of the order of Rodentia inferred from structural analysis of short retrosposon B1. *Russian Journal of Genetics,* 43, 757–768.

Verzi DH, Olivares AI, Morgan CC and Álvarez A, 2016. Contrasting phylogenetic and diversity patterns in octodontoid rodents and a new definition of the family Abrocomidae. *Journal of Mammal Evolution*, 23, 93–115.

Vianey-Liaud M, 1985. Possible evolutionary relationships among Eocene and Lower Oligocene rodents of Asia, Europe and North America, 277–309. In: Luckett WP and Hartenberger JL (eds), *Evolutionary relationships among rodents: A multidisciplinary analysis.* Plenum Press: New York.

Vianey-Liaud M and Marivaux L, 2016. Autopsie d'une radiation adaptive: phylogénie des Theridomorpha rongeurs endémiques du Paléogène d'Europe – histoire, dynamique évolutive et intérêt biochronologique. *Palaeovertebrata*, 40(3), 1–68.

Vucetich MG, Verzi DH and Hartenberger, 1999. Review and analysis of the radiation of South American Hystricognathi (Mammalia, Rodentia). *Comptes Rendus de l'Académie des Sciences - Series IIA - Earth and Planetary Science*, 365, 434–437.

Wahlert JH, 1974. The cranial foramina of protrogomorphous rodents; an anatomical and phylogenetic study. *Bulletin of the Museum of Comparative Zoology*, 146(8), 363–410.

Wang B, 1997. The Mid-Tertiary Ctenodactylidae (Rodentia, Mammalia) of eastern and central Asia. *Bulletin of the American Museum of Natural History,* 234, 1–99.

Wang GC, Cao K, Zhang KX, Wang A, Liu C, Meng YN and Xu YD, 2011. Spatio-temporal framework of tectonic uplift stages of the Tibetan Plateau in Cenozoic. *Science China Earth Science,* 54, 29–44.

Wang X, Wang Y, Li Q, Tseng ZJ, Takeuchi GT, Deng T, Xie G, Chang M and Wang N, 2015. Cenozoic vertebrate evolution and paleoenvironment in Tibetan Plateau: progress and prospects. *Gondwana Research,* 27, 1335–1354.

Wang YQ, Li CK and Li DS, 2016. A synopsis of Paleocene stratigraphy and vertebrate paleontology in the Qianshan Basin, Anhui, China. *Vertebrata PalAsiatica*, 54, 89–120.

Waterhouse GR, 1839. Observations on the Rodentia with a view to point out groups as indicated by the structure of the crania in this order of mammals. *Magazine of Natural History*, 3, 90–96.

Wheeler TJ and Kececioglu JD, 2007. Multiple alignment by aligning alignments. *Bioinformatics*, 23, 559–568.

Wible JR, Wang Y, Li C, Dawson MR, 2005. Cranial anatomy and relationships of a new ctenodactyloid (Mammalia, Rodentia) from the early Eocene of Hubei Province, China. *Annals of Carnegie Museum*, 74, 91–150.

Wible JR, Rougier GW, Novacek MJ and Asher RJ, 2007. Cretaceous eutherians and Laurasian origin for placental mammals near the K/T boundary. *Nature*, *447*(7147), 1003–1006.

Wiens JJ, 1998. Does adding characters with missing data increase or decrease phylogenetic accuracy? *Systematic Biology*, 47, 625–640.

Wiens JJ, 2003a. Incomplete taxa, incomplete characters, and phylogenetic accuracy: is there a missing data problem? *Journal of Vertebrate Paleontology*, 23, 297–310.

Wiens JJ, 2003b, Missing data, incomplete taxa, and phylogenetic accuracy. *Systematic Biology*, 52, 528–538.

Wiens JJ, 2005. Can incomplete taxa rescue phylogenetic analyses from long-branch attraction? *Systematic Biology*, 54, 731–742.

Wiens JJ and Moens DS, 2008. Missing data and the accuracy of Bayesian phylogenetics. *Journal of Systematics and Evolution*, 46(3), 307–314.

Wiens JJ, Kuczynski CA, Townsend T, Reeder TW, Mulcahy DG and Sites JW, 2010. Combining phylogenomics and fossils in higher-level squamate reptile phylogeny: molecular data change the placement of fossil taxa. *Systematic Biology,* 59(6), 674–688.

Wills MA, 1998. A phylogeny of recent and fossil Crustacea derived from morphological characters. *Systematics Association Special Volume Series,* 55, 189–209.

Wilson RW, 1935. Cricetine-like rodents from the Sespe Eocene of California. *Proceedings of the National Academy of Sciences*, 21, 26–32.

Wilson DE and Reeder DM, 2005. *Mammal species of the world volume 2.* Johns Hopkins University Press: Baltimore

Wilson DE, Latcher TE and Mittermeier RA, 2016. *Handbook of the mammals of the world: lagomorphs and rodents I*. Lynx Editions: Barcelona.

Wood AE, 1937. The mammalian fauna of the White River Oligocene Part II – Rodentia. *Transactions of the American Philosophical Society*, 28(2), 155–269.

Wood AE, 1962. The early Tertiary rodents of the family Paramyidae. *Transactions of the American Philosophical Society*, 52, 1–261.

Wood AE, 1965. Grades and clades among rodents. *Evolution*, 19, 115–130.

Wood AE, 1974. Early Tertiary vertebrate faunas, Vieja Group, Trans-Pecos Texas: Rodentia. *Texas Memorial Museum Bulletin*, 21, 1–112.

Wood AE, 1976. The Oligocene Rodents *Ischyromys* and *Titanotheriomys* and the Content of the Family Ischyromyidae, 244–277. In: Churcher CS (ed), *Essays on Palaeontology in Honour of Loris Shano Russell*. University of Toronto Press: Toronto.

Wood AE, 1977. The Rodentia as clues to Cenozoic migrations between the Americas and Europe and Africa. In: West RM (ed), *Paleontology and plate tectonics*. Milwaukee Public Museum Special Publications in Biology and Geology No. 2: Wisconsin.

Wood AE, 1980. The Oligocene rodents of North America. *Transactions of the American Philosophical Society*, 70(5), 1–68.

Wood AE and Wilson RW, 1936. A suggested nomenclature for the cusps of the cheek teeth of rodents. *Journal of Paleontology*, 10 (5), 388–391.

Wu S, Wu W, Zhang F, Ye J, Ni X, Sun J, Edwards AV, Meng J and Organ CL, 2012. Molecular and paleontological evidence for a post-Cretaceous origin of rodents. *PLoS ONE*, 7(10), e46445.

Wu X, Dong S, Liu S, Su X, Han Y, Shi J, Zhang Y, Zhao Z, Sha W, Zhang X and Gao F, 2017. Predicting the shift of threatened ungulates' habitats with climate change in Altun Mountain National Nature Reserve of the Northwestern Qinghai-Tibetan Plateau. *Climatic Change*, 142(3), 331–44.

Wyss AR, Flynn JJ, Norell MA, Swisher III CC, Charrier R, Novacek MJ and McKenna MC, 1993. South America's earliest rodent and recognition of a new interval of mammalian evolution. *Nature*, 365, 434–437.

Xiong Q, Halmy MW, Dakhil MA, Pandey B, Zhang F, Zhang L, Pan K, Li T, Sun X, Wu X and Xiao Y, 2019. Concealed truth: Modeling reveals unique Quaternary distribution dynamics and refugia of four related endemic keystone Abies taxa on the Tibetan Plateau. *Ecology and Evolution*, 9(24), 14295–316.

Xu RC and Li Q, 2020. New skulls of ctenodactyloids from the Early Oligocene of Ulantatal, Nei Mongol, China. *Vertebrata PalAsiatica*, 58(4), 305–327.

You J, Qin X, Ranjitkar S, Lougheed SC, Wang M, Zhou W, Ouyang D, Zhou Y, Xu J, Zhang W and Wang Y, 2018. Response to climate change of montane herbaceous plants in the genus Rhodiola predicted by ecological niche modelling. *Scientific Reports*, 8(1), 5879.

Yu Y, Harris AJ, Blair C and He X, 2015. RASP (reconstruct ancestral state in phylogenies): a tool for historical biogeography. *Molecular Phylogenetics and Evolution*, 87, 46–49.

Zachos J, Pagani M, Sloan L, Thomas E and Billups K, 2001. Trends, rhythms, and aberrations in global climate 65 Ma to present. *Science,* 292, 686–693.

Zachos J, Dickens GR and Zeebe RE, 2008. An early Cenozoic perspective on greenhouse warming and carbon-cycle dynamics. *Nature*, 451, 279–283.

Zhang J, Jiang F, Li G, Qin W, Wu T, Xu F, Hou Y, Song P, Cai Z and Zhang T, 2021. The four antelope species on the Qinghai-Tibet plateau face habitat loss and redistribution to higher latitudes under climate change. *Ecological Indicators*, 123, 107337.

Zhang YM, William JL and Lucky A, 2019. Understanding UCEs: a comprehensive on using ultraconserved elements for arthropod phylogenomics. *Molecular Phylogenetics, Phylogenomics and Phylogeography*, 3(5), 1–12.

Zittel KA von, 1893. *Handbuch der Palaeontologie, Sect. I. Palaeozoologie, Vol. IV, Vertebrata (Mammalia).* R. Oldenbourg: Munich.

Appendix 1

All appendices, except item 1c, can be accessed through the USB stick accompanying the hard copy of this thesis or via Apollo, the University of Cambridge Repository, where the electronic copy of this thesis is archived.

Appendix item	Filename	Description	Location
1a	"Specimens and literature consulted for morphology coding.xlsx"	Spreadsheet of specimens coded for morphology in this study plus changes to Asher et al's (2019) coding. Also contains a list of literature consulted when making coding decisions.	Appendices/Appendix 1
1b	"Molecular data sampled.xlsx"	Spreadsheet of sequences included in my matrix taken from Asher et al (2019) as well as the sequences newly added for *Cricetulus.*	Appendices/Appendix 1
1c	" Project 3341: glires2017"	Matrix containing the morphology coded for all of the subject and template taxa.	Accessible via MorphoBank: https://morphobank.org/index.php/home/Index
1d	"Real templates generator script.txt"	An R script adapted from Asher and Smith (in prep) to make batch files for TNT. This batch file makes topologies using real template artificial fossils. R commands to calculate the Q metric and other meta information of fossil templates are also included.	Appendices/Appendix 1
1e	"Random template generator script.txt"	An R script adapted from Asher and Smith (in prep) to make batch files for TNT. This batch file makes topologies using random template artificial fossils.	Appendices/Appendix 1
1f	"Random states generator script.txt"	An R script adapted from Asher and Smith (in prep) to make batch files for TNT. This batch file makes topologies using random states artificial fossils.	Appendices/Appendix 1
1g	"Consensus trees and distance metrics calculator.txt"	A script adapted from Asher and Smith (in prep) to calculate consensus trees from TNT .out files and calculate various congruence metrics with the well-corroborated tree; shared splits, corrected Robinson-Foulds and quartet divergence.	Appendices/Appendix 1
1h	"Node-to-root distance calculator.txt"	An R script adapted from Asher and Smith (in prep) that extracts the most parsimonious topologies from a TNT .out file and then calculates the node-to-root distance of each artificial fossil. The script includes R commands which calculate the difference in node-to-root distance between living taxa in the reference topology with their artificial fossil counterparts.	Appendices/Appendix 1
1i	"Reference topologies.nwk"	Two fully bifurcating topologies used in calcuations of node-to-root distances.	Appendices/Appendix 1
1j	"Artificial extinction results.xlsx"	Spreadsheet containing fossil template meta data and congruence metrics for real templates, random templates and random states.	Appendices/Appendix 1

Appendix 2

All appendices can be accessed through the USB stick accompanying the hard copy of this thesis or via Apollo, the University of Cambridge Repository, where the electronic copy of this thesis is archived.

Appendix item	Filename	Description	Location
2a	"TNT Glires matrix.tnt"	Matrix of morphological characters for extant Glires.	Appendices/Appendix 2
2b	"Optimal tree.png"	Tree based on all morphological partitions within the Glires matrix.	Appendices/Appendix 2

Appendix 3

All appendices can be accessed through the USB stick accompanying the hard copy of this thesis or via Apollo, the University of Cambridge Repository, where the electronic copy of this thesis is archived.

Appendix item	Filename	Description	Location
3a	'MrBayes alignment and commands.txt'	MrBayes batch file including morph+coding alignment. MrBayes commands, which includes partitioning scheme and models used.	Appendices/Appendix 3
3b	'RAxML alignments and commands.phy'	Phylip formatted alignment of my UCEs dataset with commands given to RAxML included at the bottom.	Appendices/Appendix 3
3c	'TreePL config file.txt'	Configuration file given to TreePL, which includes node calibrations and parameter commands.	Appendices/Appendix 3
3d	'Extrapolated squares calculator.txt'	A text file containing R commands to calculate extrapolated squares estimates of generic diversity.	Appendices/Appendix 3
3e	'Glires occurrence data.csv'	A spreadsheet containing cleaned occurrence records of Glires diversity.	Appendices/Appendix 3
3f	'Veizer and Zachos.csv'	$\delta^{18}O$ data from Veizer (1999) and Zachos et al. (2001).	Appendices/Appendix 3
3g	'MrBayes Glires tree with isotope data.png'	An extended version of Figure 19, which includes non-glires taxa.	Appendices/Appendix 3
3h	'RAxML Glires tree with isotope data.png'	An extended version of Figure 20, which includes non-glires taxa	Appendices/Appendix 3
3i	'MrBayes Glires tree with ancestral ranges.png'	An extended version of Figure 21, which includes non-glires taxa.	Appendices/Appendix 3
3j	'RAxML Glires tree with ancestral ranges.png'	An extended version of Figure 22, which includes non-glires taxa.	Appendices/Appendix 3

Appendix 4

All appendices can be accessed through the USB stick accompanying the hard copy of this thesis or via Apollo, the University of Cambridge Repository, where the electronic copy of this thesis is archived.

Appendix item	Filename	Description	Location
4a	'Specimens microCT scanned.xlsx'	Information on specimens microCT scanned, including specimen numbers and masseter types.	Appendices/Appendix 4
4b	'Landmarks.xlsx'	Description of cranial 3D landmarks adapted from Morris et al. (2018).	Appendices/Appendix 4
4c	'Eigenvalues and shape variation.xlsx'	Eigenvalues and shape variation explained by each principal component calculated by the geometric morphometric analyses.	Appendices/Appendix 4