Universität Potsdam

Dissertation

On the molecular evolution of sengis (Macroscelidea)

Justus Hagemann

Zur Erlangung des akademischen Grades
"doctor rerum naturalium"
(Dr. rer. nat.)
In der Wissenschaftsdisziplin "Vergleichende Genomik"

eingereicht an der

Mathematisch-Naturwissenschaftlichen Fakultät
Institut für Biochemie und Biologie
der Universiät Potsdam

June 17, 2024

Autor:
Justus Hagemann

Hauptbetreuer und Gutachter:
Prof. Dr. Michael Hofreiter
Zweitbetreuer und Gutachter:
PD. Dr. Faysal Bibi
Weiterer Gutachter
Prof. Dr. Mark S. Springer

"A small step for me, a gigantic leap for mankind"

improved from *Neil Amstrong*

Declaration of Authorship

I, Justus Hagemann, hereby declare that the work presented within this thesis titled, **"On the molecular evolution of sengis (Macroscelidea)"**, is purely my own unless stated otherwise. I furthermore hereby declare that this thesis has not been submitted to any university yet.

Chapter 4:

"Is it inappropriate to ask for your age? Evaluating parameter impact on tree dating in a challenging clade (Macroscelidea)". This manuscript was published in "Molecular Phylogenetics and Evolution" in June, 2023. It is openly accessible online at https://doi.org/10.1016/j.ympev.2023.107756. I carried out most of the laboratory work, all bioinformatic processing of sequence information and analysis, produced all figures and wrote the first version and all subsequent ones of the manuscript. Silke Abelt and Michaela Preick helped me with the laboratory work. Patrick Arnold organized the samples. All authors contributed their knowledge and thoughts to the manuscript.

Chapter 5:

"Four-toed Sengi (*Petrodromus tetradactylus*, Afrotheria, Mammalia) museomics reveal a crucial role of East African forests in macroscelidean diversification". This manuscript was submitted to "Zoological Journal of the Linnean Society" on the 25th of August 2023 and has been under review since. I carried out most of the laboratory work with support from Silke Abelt, Michaela Preick and Luis Victoria Nogales. Luis Victoria Nogales produced the haplotype networks. Patrick Arnold organized the samples, conducted the biogeographic modelling and contributed to the the figures capturing the modelling results. I did all other analysis and bioinformatic processing of sequence information, produced all figures and maps and wrote the first version and all subsequent ones of the manuscript. All authors contributed their knowledge and thoughts to the manuscript.

Chapter 6:

"A genomic approach on the enigmatic evolution of sengis (Macroscelidea)". This manuscript is in preparation for submission to the scientific journal "Molecular Biology and Evolution". Michaela Preick, Silke Abelt and me conducted the laboratory work. Patrick Arnold organized the samples. I assembled and annotated all genomes. The analysis, figures and written method section of the manuscript on the olfactory receptor proteins was contributed by Rohan Nath and Arunkumar Krishnan. I conducted all other bioinformatic processing and analysis of the genomic information, produced all other figures, wrote the first version of the manuscript and all subsequent ones. All authors contributed their knowledge and thoughts to the manuscript.

Justus Hagemann:

_____

Date:

_____

Contents

x

xii

List of Abbreviations

CFP   Crown fossil prior

DNA   Deoxyribonucleic acid

DRC   Democratic Republic of the Congo

HPD   Highest posterior density

ILS   Incomplete lineage sorting

KY   Thousand years

MRCA   Most recent common Ancestor

MYA   Million years ago

NGS   Next Generation Sequencing

PCR   Polymerase Chain Reaction

Chapter 1

Summary

This thesis focuses on the molecular evolution of Macroscelidea, commonly referred to as sengis. Sengis are a mammalian order belonging to the Afrotherians, one of the four major clades of placental mammals. Sengis currently consist of twenty extant species, all of which are endemic to the African continent. They can be separated in two families, the soft-furred sengis (Macroscelididae) and the giant sengis (Rhynchocyonidae). While giant sengis can be exclusively found in forest habitats, the different soft-furred sengi species dwell in a broad range of habitats, from tropical rain-forests to rocky deserts.

Our knowledge on the evolutionary history of sengis is largely incomplete. The high level of superficial morphological resemblance among different sengi species (especially the soft-furred sengis) has for example led to misinterpretations of phylogenetic relationships, based on morphological characters. With the rise of DNA based taxonomic inferences, multiple new genera were defined and new species described. Yet, no full taxon molecular phylogeny exists, hampering the answering of basic taxonomic questions. This lack of knowledge can be to some extent attributed to the limited availability of fresh-tissue samples for DNA extraction. The broad African distribution, partly in political unstable regions and low population densities complicate contemporary sampling approaches. Furthermore, the DNA information available usually covers only short stretches of the mitochondrial genome and thus a single genetic locus with limited informational content.

Developments in DNA extraction and library protocols nowadays offer the opportunity to access DNA from museum specimens, collected over the past centuries and stored in natural history museums throughout the world. Thus, the difficulties in fresh-sample acquisition for molecular biological studies can be overcome by the application of museomics, the research field which emerged from those laboratory developments.

This thesis uses fresh-tissue samples as well as a vast collection museum specimens to investigate multiple aspects about the macroscelidean evolutionary history.

Chapter 4 of this thesis focuses on the phylogenetic relationships of all currently

known sengi species. By accessing DNA information from museum specimens in combination of fresh tissue samples and publicly available genetic resources it produces the first full taxon molecular phylogeny of sengis. It confirms the monophyly of the genus *Elephantulus* and discovers multiple deeply divergent lineages within different species, highlighting the need for species specific approaches. The study furthermore focuses on the evolutionary time frame of sengis by evaluating the impact of commonly varied parameters on tree dating. The results of the study show, that the mitochondrial information used in previous studies to temporal calibrate the Macroscelidean phylogeny led to an overestimation of node ages within sengis. Especially soft-furred sengis are thus much younger than previously assumed. The refined knowledge of nodes ages within sengis offer the opportunity to link e.g. speciation events to environmental changes.

Chapter 5 focuses on the genus *Petrodromus* with its single representative *Petrodromus tetradactylus*. It again exploits the opportunities of museomics and gathers a comprehensive, multi-locus genetic dataset of *P. tetradactylus* individuals, distributed across most the known range of this species. It reveals multiple deeply divergent lineages within *Petrodromus*, whereby some could possibly be associated to previously described sub-species, at least one was formerly unknown. It underscores the necessity for a revision of the genus *Petrodromus* through the integration of both molecular and morphological evidence. The study, furthermore identifies changing forest distributions through climatic oscillations as main factor shaping the genetic structure of *Petrodromus*.

Chapter 6 uses fresh tissue samples to extent the genomic resources of sengis by thirteen new nuclear genomes, of which two were de-novo assembled. An extensive dataset of more than 8000 protein coding one-to-one orthologs allows to further refine and confirm the temporal time frame of sengi evolution found in Chapter 4. This study moreover investigates the role of gene-flow and incomplete lineage sorting (ILS) in sengi evolution. In addition it identifies clade specific genes of possible outstanding evolutionary importance and links them to potential phenotypic traits affected. A closer investigation of olfactory receptor proteins reveals clade specific differences. A comparison of the demographic past of sengis to other small African mammals does not reveal a sengi specific pattern.

Chapter 2

Zusammenfassung

Diese Dissertation untersucht die molekulare Evolution von Macroscelidea, auch als Sengis oder Rüsselspringer bezeichnet. Sengis sind eine Ordnung der Afrotheria, einer der vier Hauptkladen der plazentalen Säugetiere. Aktuell gibt es zwanzig beschriebene Sengiarten, die alle ausschließlich auf dem afrikanischen Kontinent vorkommen. Sengis können in zwei Familien unterteilt werden: die Elephantenspitzmäuse zusammen mit den Rüsselratten bilden die Macroscelididae und die Rüsselhündchen die Rhynchocyonidae. Während Rhynchocyonidae ausschließlich in Waldhabitaten zu finden sind, bewohnen verschiedene Macroscelididaearten ein breites Spektrum von Lebensräumen, von tropischen Regenwäldern bis zu felsigen Wüsten.

Unser Wissen über die evolutionäre Geschichte der Sengis ist äußerst unvollständig. Der hohe Grad an morphologischer Ähnlichkeit zwischen verschiedenen Sengiarten (insbesondere innerhalb der Macroscelididae) hat beispielsweise zu Fehlinterpretationen phylogenetischer Beziehungen auf der Grundlage morphologischer Merkmale geführt. Mit dem Aufkommen DNA-basierter taxonomischer Forschung wurden mehrere neue Gattungen definiert und neue Arten beschrieben. Dennoch existiert derzeit keine vollständige molekulare Phylogenie, was die Beantwortung grundlegender taxonomischer Fragen und tiefergehende evolutionsbiologische Analysen erschwert. Dieser Mangel an Wissen kann zum Teil auf die begrenzte Verfügbarkeit von frischen Gewebeproben für die DNA-Extraktion zurückgeführt werden. Die weite Verbreitung in Afrika, teilweise in politisch instabilen Regionen und geringe Populationssdichten von Sengis erschweren das Sammeln von frischem Probenmaterial, was für die Extraktion von DNA genutzt werden kann. Darüber hinaus deckt die bis jetzt verfügbare DNA-Information über Sengis häufig nur kurze Abschnitte des mitochondrialen Genoms ab und damit einen einzelnen genetischen Lokus mit begrenztem Informationsgehalt.

Fortentwicklungen von DNA-Extraktions-Protokollen und Library-Protokollen bieten heutzutage die Möglichkeit, auf DNA von Museumsexemplaren zuzugreifen, die über die letzten Jahrhunderte gesammelt und in Naturkundemuseen weltweit

aufbewahrt werden. Somit können die Schwierigkeiten bei der Beschaffung von Frischproben für molekularbiologische Studien überwunden werden.

Diese Dissertation verwendet sowohl Frischgewebeproben als auch eine umfangreiche Sammlung von Museumssproben, um verschiedene Aspekte der evolutionären Geschichte der Sengis molekularbiologisch zu untersuchen.

Kapitel 4 dieser Dissertation konzentriert sich auf die phylogenetischen Beziehungen aller derzeit bekannten Sengiarten. Durch das Generieren von DNA-Information aus Museumsexemplaren in Kombination mit Frischgewebeproben und öffentlich verfügbaren genetischen Ressourcen wird die erste vollständige molekulare Phylogenie aller Rüsselspringer erzeugt. Die Studie bestätigt die Monophylie der Gattung *Elephantulus* und entdeckt mehrere tief divergente Linien innerhalb verschiedener Arten, was die Notwendigkeit speziesbezogener Ansätze verdeutlicht. Die Studie konzentriert sich außerdem auf den Zeitrahmen der Sengi-Evolution, indem sie die Auswirkungen häufig variierter Parameter auf die Datierung von Stammbäumen untersucht. Die Ergebnisse zeigen, dass die mitochondriale Information, die in früheren Studien zur zeitlichen Kalibrierung der Macroscelidean-Phylogenie verwendet wurde, zu einer Überschätzung des Alters von Arttrennungen innerhalb der Rüsselspringer geführt hat. Insbesondere die Macroscelididae sind daher viel jünger als zuvor angenommen. Das präzisere Wissen über das evolutionäre Alter von Rüsselspringern bietet die Möglichkeit, beispielsweise Artaufspaltungen mit Umweltveränderungen zu verknüpfen.

Kapitel 5 konzentriert sich auf die Gattung *Petrodromus* mit ihrem einzigen Vertreter *Petrodromus tetradactylus*. Es nutzt erneut die Museomics und sammelt einen umfassenden, genetischen Datensatz von *P. tetradactylus*-Individuen, die über den größten Teil des bekannten Verbreitungsgebiets dieser Art verteilt sind. Es zeigt mehrere tief divergente Linien innerhalb von *Petrodromus* auf, wobei einige mit zuvor beschriebenen Unterarten in Verbindung gebracht werden könnten, mindestens eine aber zuvor unbekannt war. Die Ergebnisse verdeutlichen die Notwendigkeit einer taxonomischen Überarbeitung der Gattung *Petrodromus* durch das Zusammenführen sowohl molekularer als auch morphologischer Indizien. Die Studie identifiziert außerdem sich ändernde Waldverteilungen durch klimatische Schwankungen als Hauptfaktor, der die genetische Struktur von *Petrodromus* formt.

Kapitel 6 verwendet Frischgewebeproben, um die genomischen Ressourcen der Rüsselspringer durch dreizehn neue nukleare Genome zu erweitern, von denen zwei de-novo assembliert wurden. Ein umfangreicher Datensatz von mehr als 8000 protein-kodierenden 1:1-Orthologen ermöglicht es, den zeitlichen Rahmen der Rüsselspringerevolution, der in Kapitel 4 gefunden wurde, weiter zu verfeinern und zu bestätigen. Diese Studie untersucht außerdem die Rolle von Genfluss auf

die Evolution der Rüsselspringer. Darüber hinaus identifiziert sie für bestimmte Kladen spezifische Gene von möglicherweise herausragender evolutionärer Bedeutung und verknüpft diese mit potenziell betroffenen phänotypischen Merkmalen. Eine genauere Untersuchung von Geruchsrezeptorproteinen zeigt kladespezifische Unterschiede auf.

Chapter 3

Introduction

3.1 Afrotheria and the revision of the Placentalia phylogeny based on molecular evidence

Placental mammals can be separated into four major clades: Xenarthra, Euarchontoglires, Laurasiatheria, and Afrotherians (e.g. Amrine-Madsen et al., 2003; Murphy et al., 2001a,b; Nikolaev et al., 2007; Nishihara et al., 2005; Scally et al., 2001; Springer and Murphy, 2007; Springer et al., 2004; Wildman et al., 2007). However, the knowledge of this phylogenetic grouping is rather young. Before the now common use of genetic evidence to infer phylogenetic relationships between organisms, taxonomists mostly relied on morphological characters to conduct this task. Although many of the hereby inferred relationships could be confirmed with genetic evidence, especially the relationships between different higher groups of mammals were, to some extent, proven to be wrong (Novacek, 1992; Shoshani, 1986). In some cases, convergent evolution to adapt to similar environmental niches led to morphological resemblance of evolutionary distant organisms, which were then misinterpreted as taxonomic proximity. Homoplasy was therefore falsely interpreted as homology (Foley et al., 2016).

From the previously named four major groups of placental mammals, only the Xenarthra were defined based on morphology and later confirmed with DNA analysis. The definition of the other three groups resulted from a regrouping of mammalian clades based on genetic evidence (Springer et al., 2004). The revised phylogenetic relationships of placental mammals mostly fit the tectonic movement of continental plates. The separation of continents therefore facilitated the genetic isolation of these clades from each other and their subsequent evolutionary path. While Euarchontoglires, which include, for example, primates, rodents, lagomorphs, treeshrews and Laurasiatheria, comprising bats, carnivorans, pangolins, and ungulates, evolved in the Northern Hemisphere on the former supercontinent Laurasia, Xenarthra and Afrotheria are assumed to have their origin in the Southern Hemisphere. Xenarthra

(armadillos, anteaters, tree sloths) are thought to have originated in South America while Afrotheria, as the name indicates, have an evolutionary history closely entangled with the African continent (Springer, 2022; Springer et al., 2004).

Extant Afrotherians consist of six orders, grouped into two main clades: the Paenungulata and the Afroinsectiphilia. Elephants, sea-cows, and hyraxes make up Paenungulata, whereby the exact relationships among them are still a topic of ongoing research (Springer, 2022). Afroinsectiphilia, also defined based on molecular evidence, include aardvarks, sengis, and Afrosoricida. The latter order can again be split into the three families: golden moles, African otter shrews, and tenrecs.

Afrotheria is an excellent example of genetic analysis revealing relationships overseen or misinterpreted by morphological-only-based approaches. The taxonomic entity Afrotherians, sometimes referred to as a superorder (e.g. Jennings and Rathbun, 2001), was unknown until the 1990s (Springer et al., 1997; Stanhope et al., 1998) although first indications of its existence were found already in 1977 (Jong et al., 1977) and kept accumulating (De Jong et al., 1981). The six Afrotherian orders share very little morphological resemblance, while some display a high level of morphological convergence and thus morphological similarities to other mammalian groups, which led to their phylogenetic misplacing in the past.

Tenrecs show similar adaptive traits as hedgehogs of the Laurasiatheria. The same is true for golden moles compared to the Laurasiatherian true moles, and Aardvark, at least to some extent, to the Xenarthrian anteaters. As the name already indicates, African otter shrews show high similarities to otters (Laurasiatheria), while hyraxes show similarities to Marmots (Euarchontoglires) (see e.g. Springer, 2022 for visual comparison). The extinct afrotherian Embrithopoda have stunning resemblances to extant Rhinos (Laurasiatheria).

Even with the knowledge of Afrotherian monophyly and interrelationships from molecular biological studies, no clear known morphological feature unites all Afrotherians. Multiple morphological traits have been suggested as afrotherian synapomorphies, like testicondy and a "mobile proboscis," but rejected after closer examination (Tabuce et al., 2008). A high number of Thoracolumbar vertebrae (Sánchez-Villagra et al., 2007), osteological characters of the astragalus (Seiffert, 2007; Tabuce et al., 2007), and a relative late eruption of permanent dentition (Asher and Lehmann, 2008) are considered the clearest morphological features uniting this group today.

The work presented here focuses on the molecular evolution of Macroscelidea or sengis, an Afrotherian order which shows some convergent features to other small mammals and was therefore long assumed to be part of the monophyletic group Insectivora with e.g. Scandentia (tree shrews) and Soricidae (shrews) (e.g. Novacek, 1992; Shoshani and McKenna, 1998).

3.2 Macroscelidea

Until the end of the 20th century, sengis (Macroscelidea, Butler, 1956) were referred to as elephant-shrews due to their trunk-like nose and their shrew-like overall appearance. However, this name was changed to "sengi" for two major reasons: the misleading taxonomic indication of "shrew" and the developed preference in science to call endemic organisms by their local names (Kingdon, 1997; Olbricht and Sliwa, 2014; Rathbun and Woodall, 2002; White and Ansell, 1966). "Sanje" is the Kigiriama word for sengi, a language belonging to the Mijikenda group of languages, which again is a Bantu dialect (Heine and Möhling, 1980). "Isengi" is the Lunda word for sengi, another Bantu language spoken in Zambia, Angola, and to some extent in the Democratic Republic of the Congo (DRC) (Kawasha, 2003; Olbricht and Sliwa, 2014).

(A) *R. petersi* (B) *P. tetradactylus* (C) *E. intufi*
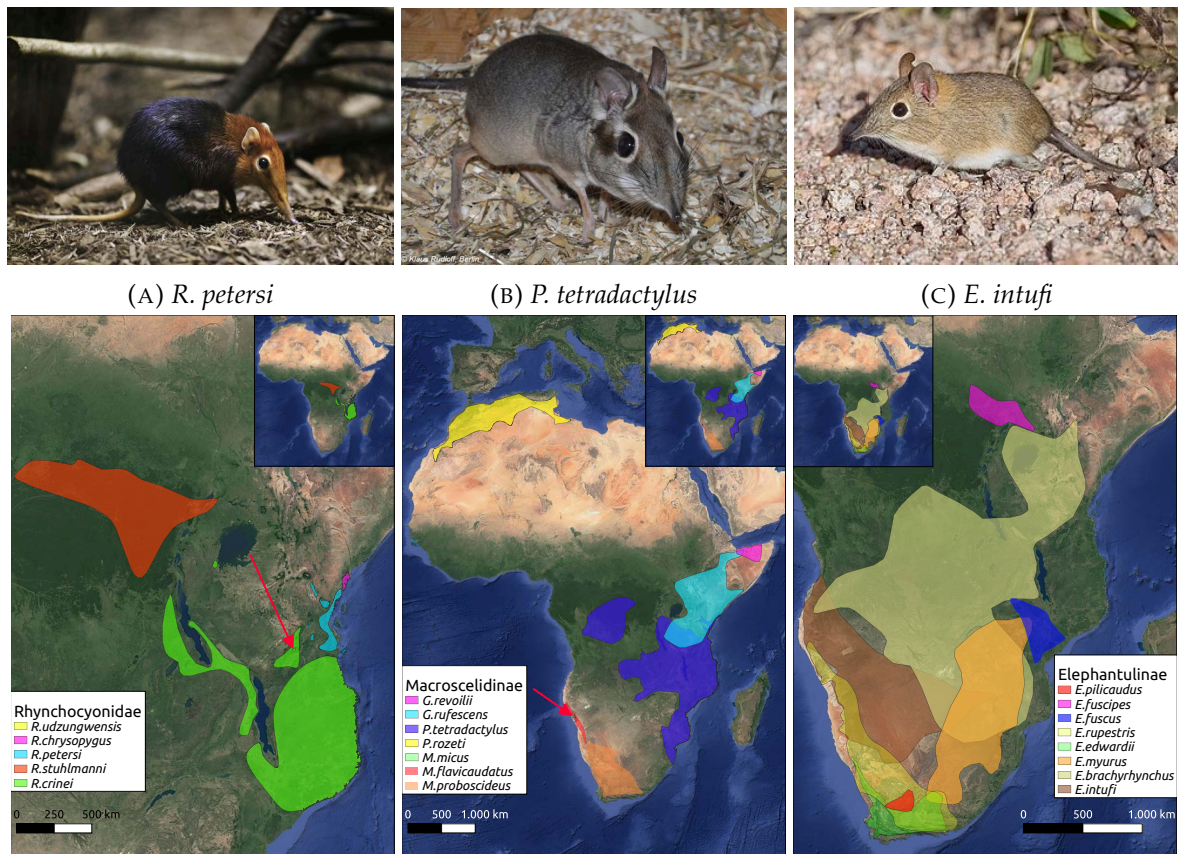
(D) Distribution of Rhynchocyonidae. Red arrow indicates the distribution of *R. udzungwensis*.

(E) Distribution of Macrosclidinae. Red arrow indicates the distribution of *M. micus*.

(F) Distribution of Elephantulinae.

FIGURE 3.1: The three mein clades of Macroscelidea, (Left) Rhynchocyonidae, (Middle) Macroscelidinae, (Right) Elephantulinae. Top row: One representative of each clade. Bottom row: Distribution of the members of the respective clade. (Sources: Wikimedia, IUCN Redlist, Google maps).

Sengis currently consist of 20 extant species, which can be separated into two phylogenetic distinct groups: the soft-furred sengis and the giant sengis. Until recently, these two groups were classified as subfamilies, Macroscelidinae and Rhynchocyoninae, respectively. However, based on morphological differences in the fossil record and in comparison to morphological differences between other mammalian families, it was argued that the taxonomic status of family would be more appropriate for giant and soft-furred sengis (Senut and Pickford, 2021). The here presented work will follow this argumentation and classifies giant sengis and soft-furred sengis as families, hence as Rhynchocyonidae and Macroscelididae throughout.

The unique combination of morphological, behavioral, reproductive, and physiological traits displayed within sengis constitutes their attractiveness for for evolutionary studies. The research record on sengis is quit extensive. An assessment of sengi literature in 2002 came up with 750 studies including the group (Rathbun and Woodall, 2002). Despite this wealth of scientific literature on sengis, our understanding of their evolution is not only surprisingly incomplete but appears to some extent more conflicting than convincing.

Life-history traits of sengis have been described as a combination of features from small antelopes and anteaters (Rathbun, 2009b). Sengis are of small size and weight (<1kg), have trunk-like mobile snouts and long flexible tongues, relatively large eyes, long tails, and predominantly feed on invertebrates. Sengis are considered the only mammalian order of which all species are (socially) monogamous. However, it was recently questioned if this assumption can be drawn from the existing behavioral data. Olivier et al., 2022 argues that conflicting evidence exists, and further studies need to be conducted in order to confirm or reject the claim of monogamy in sengis. Nonetheless, males do not seem to contribute (at least directly) to the upbringing of offspring, while maternal contact to neonates is also minimized to a single interaction for lactation per day. In contrast to soft-furred sengis, which have exposed sheltering habits, giant sengis build primitive nests out of leaf litter and give birth to non-precocial neonates, which stay inside the nest for two to three weeks after birth (Heritage, 2018; Perrin and Rathbun, 2013). Neonates of soft-furred sengis are highly precocial (Rathbun, 2009a). In contrast to many other mammals of similar size, sengis have small litter sizes of one to three, mostly two individuals (Birney and Baird, 1985). Some sengis maintain paths within their territories, which they clear from organic and inorganic debris and use them for foraging and as escape routes. Their extended hind limbs allow a swift saltatorial gait and a generally highly cursorial locomotion (Rathbun, 2009b).

Giant sengis consist of a single genus *Rhynchocyon* and currently five species, while

soft-furred sengis are more species-rich with currently fifteen species. These can be separated into two clades: the Elephantulinae, comprising the single genus *Elephantulus*, and Macroscelidinae consisting of the four genera *Macroscelides*, *Petrodromus*, *Galegeeska*, and *Petrosaltator*. The latter two were only recently defined as it was shown that the species *G. revoilii*, *G. rufescens*, and *P. rozeti* do not belong to the genus *Elephantulus* as previously assumed (Dumbacher et al., 2016; Heritage et al., 2020; Krásová et al., 2021).

In contrast to the extensive morphological differences between the six afrotherian orders, sengis show a high level of superficial resemblance, and only few morphological traits can be used to differentiate the species, especially within soft-furred sengis. This situation led to a confusing taxonomy in the past (Corbet and Hanks, 1968; Corbet and Neal, 1965). While 40 species were described by 1939, this number was reduced to fourteen after a major taxonomic revision by Corbet and Hanks in 1968. Over the past two decades, this number has again increased to 20, primarily based on genetic evidence (Carlen et al., 2017; Douady, 2001; Dumbacher et al., 2016, 2012; Heritage et al., 2020; Rovero et al., 2008; Smit, 2008; Smit et al., 2011).

Current research on sengis is hampered by the limited availability of genetic information. The accessible information is mostly restricted to a few marker genes, mostly mitochondrial ones, and does not cover all extant species. Hence, no full taxon phylogeny of sengis based on molecular evidence exists until now. Among other uncertainties, it is therefore unknown if the genus *Elephantulus*, after the recent assignments of *P. rozeti*, *G. revoilii*, and *G. rufescens* from *Elephantulus* to their current genera, is now monophyletic or if further rearrangements are required.

Besides these unresolved basic taxonomic relationships among sengis, their evolutionary time-frame is also largely unknown, as conflicting node age estimates can be found in literature (Heritage et al., 2020; Krásová et al., 2021). Sengis are assumed to have been evolving with minor or no competition from other small mammals, as Africa was isolated from other land masses from roughly 80mya to 21mya (Pekar and Deconto, 2006; Van Couvering and Delson, 2020). Based on the fossil record, the split between sengis and the next close afrotherian clade, the Afrosoricida, is estimated to be about 56my old (Gheerbrant et al., 1998; Seiffert, 2010). The evolutionary lineage of Macroscelidea is therefore assumed to be considerably old. Despite this age, the fossil record documents very little changes in sengi morphology over the past 40mya. Only the transition from a herbivorous diet to a predominantly myrmecophage one is well documented through changes in teeth morphology and a working caecum in extant species (Holroyd, 2010; Rathbun, 2009b; Woodall and Mackie, 1987). Large uncertainty exists on the estimated node ages within sengis. Depending on the study and taking error margins into account, the age of the split between soft-furred sengis

and giant sengis is estimated to be between 22.1 and 39.3 million years old (Heritage et al., 2020; Krásová et al., 2021). The uncertainty around the age of the crown node of soft-furred sengis is even more pronounced with 7.4-31.3 million years (Heritage et al., 2020; Krásová et al., 2021). Obviously, this uncertainty in node ages hinders our ability to connect ecological, environmental, and geological events to radiations or speciation events of the macroscelidean phylogeny. Hypotheses on which of these forces shaped the evolution of sengis are therefore of rather speculative nature.

Although the formation of the Tauride land bridge between Africa and Eurasia enabled extensive Afro-Laurasian faunal exchange (Pekar and Deconto, 2006; Van Couvering and Delson, 2020), effects of these events on the evolution of sengis are mostly unclear. The persistence of extant sengi species in the wild serves as evidence of their evolutionary success in maintaining their ecological niches amidst competition from other small mammals. This aspect suggests the existence of adaptations which facilitate this ability. However, the latter facet of sengi evolution, as well as the broad differences in habitat usage by different sengi species, are in stark contrast to the circumstance that no sengi species has been recorded in sub-Saharan western Africa, nor does any evidence exist that sengis ever dispersed out of Africa. New fossil findings (Stevens et al., 2022) and additional genetic data can help to clarify the time-frame of sengi evolution. This knowledge is essential for a better understanding of sengi evolution. However, sengis have very low population densities of 1-2 individuals per hectare (Oxenham and Perrin, 2009). Their broad African distribution, also in politically unstable regions, complicates contemporary sampling of tissue material for DNA analysis. Utilizing sengi specimens, collected over the past centuries and stored in natural history museums throughout the world, poses the possibility to circumvent time-consuming and resource-intensive sampling while generating additional genetic data of sengis. This information will help to address standing questions about the macroscelidean evolution outlined above in order to increase our scientific understanding of this remarkable group of mammals.

3.3 Museomics

Around the 16th century, naturalists started to collect and archive biodiversity in private and natural history museums' collections (Raxworthy and Smith, 2021). These, since then constantly extended collections, constitute a research resource of immense and versatile value for natural science. They represent a spatiotemporal record of the global biodiversity over considerable time frames, allowing to address a broad variety of research questions.

Before the rise of DNA-based analysis of evolution, research on these materials mostly focused on morphological studies. Although certainly of large value, specimens of museum collections can also pose certain difficulties for morphological studies. Some traits, like, for example, the skin color of aquatic organisms, oftentimes cannot be preserved over time, complicating the comparison of freshly sampled material to museum specimens. It is also common to not conserve the whole specimen but, for example, only the skin for small mammals. Thus, taxonomic comparison between contemporary samples and museum samples can be complicated if it relies on morphological traits absent in the museum sample. Frequently the holotype of a species is such a museum specimen, prohibiting the accurate taxonomic definition of whole species complexes (Agne et al., 2022a).

Nowadays, the advances in DNA sequencing technology in conjunction with extraction and library protocols tailored to meet the challenges of poorly preserved samples (further described below) allow the exploitation of genetic information from these museum collections, assessed over the past centuries. Museomics, the research field which emerged from these developments, uses museum samples to obtain and analyses genomic or proteomic information for molecular biological studies. It allows addressing research questions which were previously impossible to answer, like the genetic analysis of extinct species (e.g. Springer et al., 2015), the investigation of demographic developments of groups of organisms over long temporal periods (Valk et al., 2019) and to clarify taxonomic questions by accessing the genetic profiles of type specimens for comparison (Agne et al., 2022a). In addition, museomics offers the opportunity to perform most research possible with DNA from modern samples, like inferring aspects about a species' biogeography, resolving phylogenetic questions, etc. Furthermore, museum collections not only harbor extinct species, but often also species of which contemporary sampling is complicated by, e.g., the political instability in a region, the low population densities requiring vast, expensive and time-consuming sampling efforts, or low population sizes prohibiting further removal of individuals. Museomics, therefore, established itself as a valuable tool to study evolutionary aspects of basically any kind of organism, which is part of the

natural history collections around the world.

Challenges of extracting and analyzing DNA from museum samples, referred to as historical or archival DNA, sometimes as hDNA, are similar, although mostly not as pronounced, as those of ancient DNA. No sharp definition separating the two terms historical and ancient DNA exists. In scientific literature, historical DNA usually refers to DNA from samples which were stored under conditions not optimized for DNA preservation and are roughly between 30-200 years old (Raxworthy and Smith, 2021). DNA from older samples is usually called ancient DNA. "Not optimized for DNA preservation" means that the samples usually are stored at room temperature, as dry samples or within fluid and not, e.g., frozen which decelerates the decay of DNA after the death of an organism. DNA quality from museum samples can vary considerably across different samples and different collections. The long storage times under unfavorable conditions usually result in low overall DNA content and fragmentation of the DNA, although mostly not as severe as the damages of ancient DNA. Thus, the application of laboratory protocols optimized for historical/ancient DNA needs to be applied in order to access sufficient amounts of DNA for further analysis (Hofreiter et al., 2001).

Early DNA extraction methods had the offset that they required the destruction of the biological material which was used for DNA extraction, putting collection curators in the difficult position to either allow at least partial damage to the irreplaceable specimen for a chance of scientific insights, or prevent sampling in order to leave the collection item intact. However, methodology has also progressed in terms of minimizing the damage to the valuable specimen (Paijmans et al., 2020; Rohland et al., 2004).

3.3.1   Challenges of historical DNA and how to address them

Following the death of an organism, the degradation of its organic components starts immediately, including the degradation of its genetic material. These processes can be slowed down by extracting the DNA in order to separate it from (bio-)chemical agents which drive the degradation, like DNases and/or by storing the material at as low temperatures as possible. Fossils or museum samples are usually not stored under such conditions, resulting in relatively low DNA concentrations. The DNA molecules present are fragmented, whereby the length of these fragments might be to some extent, driven by the age of the specimen (Zimmermann et al., 2008). Additional DNA damage patterns like single-strand breaks and cytosine deamination further complicate the access to the genetic material as well as the

common contamination with modern DNA from, e.g., microbes living on the specimen, humans handling it, or storage in close proximity to other DNA-containing organic material. Common protocols used for modern samples usually fail or yield only very low amounts of genetic information. However, methodological advances to meet the challenges of historical/ancient DNA have been developed, paving the path to access and analyze historical DNA (Dabney et al., 2013; Gansauge et al., 2017; Horn, 2012; Maricic et al., 2010). As the here presented work vastly utilized museum specimens, I briefly describe two approaches to increase DNA yield from historical samples. For more detailed information, please see Chapter 4 and Chapter 5.

Single stranded libraries

After extraction, DNA molecules need to be built into libraries in order to make them accessible for next-generation sequencing (NGS) approaches like Illumina sequencing. Library protocols optimized for modern DNA commonly ligate double-stranded adapter molecules to both ends of the double-stranded target DNA molecules, which can then be used for amplification via polymerase chain reaction (PCR) and subsequent sequencing. However, if single-stranded breaks, meaning missing bases in one or both strands of the double helix, exist, these molecules break apart during the denaturation step of the PCR. Amplification via PCR is then impossible, as one or both adapters, and thus primer binding sites, are missing. Hence, if double-stranded library preparation protocols are applied to extracts of historical DNA, the risk of loss of the already limited amount of information (total endogenous DNA amount) is high. In brief, single-stranded library protocols first denature the double-stranded target DNA and subsequently ligate adapter molecules to these molecules, which can then be further processed and amplified for NGS. This way, the loss of historical (target) DNA is efficiently reduced (Dabney et al., 2013; Gansauge et al., 2017; Gansauge and Meyer, 2013).

Hybridization target enrichment

Another challenge when working with museum specimens is the low amount of endogenous DNA, meaning DNA belonging to the target organism. Due to the degradation of DNA over time, the amount of endogenous DNA decreases and is usually only present in trace amounts (Rohland and Hofreiter, 2007). Its relative concentration is further reduced by contamination with modern DNA from other sources. Endogenous DNA content of historical samples varies, but

not unusually is below 2%. NGS techniques like Illumina shotgun sequencing result in a randomly drawn representation of DNA molecules within a sequencing library. If the amount of endogenous DNA within the library is therefore low, most sequenced reads will represent non-target DNA. This has multiple disadvantages, including ineffectiveness in terms of sequencing cost and labor, and the risk that non-target DNA is misinterpreted as target DNA, potentially falsifying the results drawn from its informational content (e.g. Pääbo, 1985). Furthermore, if interested in specific genomic regions, especially from historical and ancient samples, it requires enormous shotgun sequencing efforts in order to generate sufficient data of the respective locus. Although oftentimes there is no alternative, especially when aiming at whole genomes, approaches exist to enrich the proportion of target DNA within the DNA extract/library and to target specific regions of interest. One such method, also applied during the here presented work, is target enrichment through hybridization capture.

This method makes use of the fact that phylogenetically closer organisms also share more similarities within their genomes compared with more distant ones. In brief, hybridization capture uses bait sequences gained from a modern sample of a somewhat phylogenetically close organism or custom-synthesized DNA to which a biotin molecule is covalently bound. By heat denaturation of the DNA strands and subsequent cooling, the target DNA molecules will hybridize at least to some extent to the bait-DNA, in contrast to the non-target DNA due to its larger sequence differences. Afterwards, streptavidin-coated magnetic beads are added to the reaction. The biotin molecules of the baits bind to streptavidin, which allows washing away non-target DNA while withholding the magnetic beads and bound target DNA with a magnet. A second round of denaturation releases the target DNA from the baits for further processing. Hybridization capture is nowadays well established in genomics and works with custom-made baits from PCR products (Maricic et al., 2010) or synthesized DNA, across large phylogenetic distances and broad phylogenetic ranges (e.g. Agne et al., 2022c; Hutter et al., 2022; Li et al., 2013; Mohandesan et al., 2017; Paijmans et al., 2016). It is therefore a cost-efficient way to gain information on specific genomic regions while reducing the fraction of non-target DNA.

3.4 Thesis outline

This thesis addresses multiple questions about the phylogenetic and evolutionary past of Macroscelidea by using a broad variety of genetic and genomic approaches and DNA from museum samples as well as modern ones. Differences in data set composition facilitate the investigation of questions on different phylogenetic levels, from general Macroscelidean evolution to population genetics of a single sengi species.

Chapter 4 presents a full species phylogeny of sengis based on multi-locus DNA from mitochondrial and nuclear genes from modern and museum samples. Besides addressing standing taxonomic questions, it clarifies the time frame of sengi evolution, of which vastly diverging estimates have been published in the past, hampering the connection of sengi evolution to events such as geological or climatic changes. Besides the focus on sengis, this research is of general value for the scientific community, as it closely investigates and mostly explains the effects of commonly varied parameters in tree dating on the outcome of the temporal calibration. It will therefore be of help for dating studies on groups with a similarly incomplete fossil record on how to to choose appropriate parameters and evaluate those.

In contrast to Chapter 4, Chapter 5 focuses on the population genetics of a single species, *P. tetradactylus*. It therefore investigates sengi evolution on a different phylogenetic level. By the extensive utilization of museum specimens, it creates a comprehensive and unprecedented data set of this species across most of its known spatial distribution. By making use of the findings of Chapter 4, it provides a temporal framework for the evolution of *Petrodromus*, which is then used for biogeographic inferences about the genus' evolution. The East African forest system and its distributional changes through climatic oscillations are identified as key drivers for the current population structure of *Petrodromus*. Furthermore, multiple deeply divergent lineages are identified, of which some might be in association with previously defined subspecies, while others are undescribed. Remarkably, the study finds multiple parallels between the spatial distribution of distinct *Petrodromus* lineages and the similar forest-dwelling, but distantly related giant sengi species. The study demonstrates the necessity of a taxonomic revision of the genus by using a combination of molecular and morphological evidence.

Chapter 6 again focuses on sengis as a whole, by reconstructing and conducting a comparative analysis of whole nuclear genomes of fourteen out of the twenty known sengi species. Through linked-read sequencing, it creates two de-novo genomes of sufficient quality to be used as references for the assembly of eleven shotgun genomes. High-quality annotation of these novel genomic resources resulted in

the identification of 15,000 to 18,000 protein-coding and pseudogenized genes per species. This comprehensive data set allows the identification, comparison, and functional implications of clade-specific genes which were positively selected, experienced gene-family expansion or contraction, or were lost during the course of evolution. It also inspects the role of gene flow within Macroscelidea and produces a high-confidence species tree. Chapter 6 furthermore investigates the demographic history of sengis in comparison to other small African mammals and closely examines the evolution of genes associated with the olfactory system, which is assumed to be of importance for Macroscelidea. Besides further refining the temporal framework of sengi evolution found in Chapter 6, this study identifies multiple target genes which might have played an important role during sengi evolution and the potential phenotypic traits they act on. The genomic data produced during this thesis will be of use for future research focusing on sengis, Afrotherians, or mammals as a whole.

Chapter 4

Temporal calibration

Is it inappropriate to ask for your age?
Evaluating parameter impact on tree dating in a challenging clade (Macroscelidea)

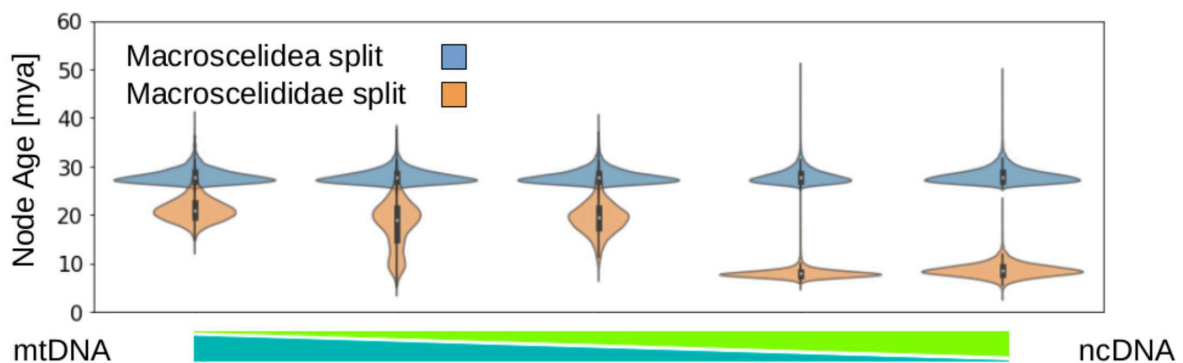**Justus Hagemann**, Michael Hofreiter, Faysal Bibi, Patricia Holroyd, Patrick Arnold

FIGURE 4.1: Graphical abstract

4.1   Abstract

Sengis (order Macroscelidea) are small mammals endemic to Africa. The taxonomy
and phylogeny of sengis has been difficult to resolve due to a lack of clear morpho-
logical apomorphies. Molecular phylogenies have already significantly revised sengi
systematics, but until now no molecular phylogeny has included all 20 extant species.
In addition, the age of origin of the sengi crown clade and the divergence age of its
two extant families remain unclear. Two recently published studies based on differ-
ent datasets and age-calibration parameters (DNA type, outgroup selection, fossil
calibration points) proposed highly different divergent age estimates and evolution-
ary scenarios. We obtained nuclear and mitochondrial DNA from mainly museum
specimens using target enrichment of single-stranded DNA libraries to generate the
first phylogeny of all extant macroscelidean species. We then explored the effects of
different parameters (type of DNA, ratio of ingroup to outgroup sampling, number
and type of fossil calibration points) and their resulting impacts on age estimates
for the origin and initial diversification of Macroscelidea. We show that, even after
correcting for substitution saturation, both using mitochondrial DNA in conjunction
with nuclear DNA or alone results in much older ages and different branch lengths
than when using nuclear DNA alone. We further show that the former effect can be
attributed to insufficient amounts of nuclear data. If multiple calibration points are
included, the age of the sengi crown group fossil prior has minimal impact on the
estimated time frame of sengi evolution. In contrast, the inclusion or exclusion of
outgroup fossil priors has a major effect on the resulting node ages. We also find that
a reduced sampling of ingroup species does not significantly affect overall age esti-
mates and that terminal specific substitution rates can serve as a means to evaluate
the biological likeliness of the produced temporal estimates. Our study demonstrates
how commonly varied parameters in temporal calibration of phylogenies affect age
estimates. Dated phylogenies should therefore always be seen in the context of the
dataset which was used to produce them.

4.2 Introduction

Macroscelidea or sengis, also called elephant-shrews, currently consist of 20 extant species. They belong to the Afrotheria, a major clade of placental mammals that remains poorly characterized morphologically (Gheerbrant et al., 2014) and has an ancient African origin. Sengis can be divided into two morphologically distinct families, the Macroscelididae (soft-furred sengis) and the Rhynchocyonidae (giant sengis), all of which are endemic to Africa. Sengis have multiple remarkable features that, in combination, distinguish them from other mammals. All have elongated, highly flexible noses and extended hindlimbs resembling those of antelopes and allowing a swift saltatorial, ungulate-like gait (Rathbun, 1979, 2009b). The species for which sufficient knowledge exists predominantly feed on insects, have a life-long monogamous lifestyle and give birth to highly precocial neonates to whom they provide minimal parental care (Heritage et al., 2020; Rathbun, 1979, 2009b; Rathbun and Rathbun, 2006). Some species maintain paths within their well-defined territory by clearing them from organic litter and small stones (Rathbun, 2009b). Despite a wide range of preferred habitats — from tropical rainforest to deserts — and a broad African distribution, congeneric species show limited morphological differentiation, which has resulted in unreliable phylogenetic reconstructions. By 1939, 40 species were described, but after a major revision by Corbet and Hanks (1968), this number was reduced to 14. The arrival of DNA sequencing led to a second major systematic revision, increasing the number of recognized species to 20, and resulting in the designation of three new genera (Carlen et al., 2017; Douady et al., 2003; Dumbacher et al., 2016; Heritage et al., 2020; Krásová et al., 2021; Rovero et al., 2008; Smit, 2008; Smit et al., 2011). Even though an increase in phylogenetic studies on macroscelideans in the past decade has improved our knowledge of this group, the rarity of some species, their low population densities and often small geographic ranges resulted in no study to date has included all extant species.

Another unresolved question is the age of origin of the sengi crown clade. Two recently published phylogenies came up with highly differing age estimates. The most basal split within Macroscelidea (here designated Node 0), which separates giant sengis (Rhychocyonidae) from soft-furred sengis (Macroscelididae), was dated to either approx. 32.7 (26.9–39.3 95% highest posterior density (HPD)) (Heritage et al., 2020) or 25.7 (22.1–30.9 HPD) million years ago (Ma) (Krásová et al., 2021). The age difference is even more pronounced for the divergence of the two major subfamilies of soft-furred sengis, Macroscelidinae and Elephantulinae (here designated Node 1), dated to approx. 28.5 (23.2–31.3 HPD) (Heritage et al., 2020) and 11.8 (7.4–16.7 HPD) million years ago (mya) (Krásová et al., 2021), respectively. These differences in age

estimates hinder our ability to reconstruct the diversification of sengis with respect to important biogeographic and climatic events.

Widely differing age estimates could be explained by differences in the data sets and parameters used. Whereas the full set of available data (mitochondrial, nuclear, outgroups, multiple fossils) was used by Heritage et al., 2020, Krásová et al., 2021 used just two nuclear genes and no outgroups. The latter authors justified their approach by arguing that the sengi fossil record currently only permits the age-calibration of the crown clade. All other possible fossil age priors are located outside Macroscelidea. The authors argued that the inclusion of outgroup age priors would falsely draw nodes to older ages, a possible outcome considering the evolutionary distance to other afrotherian groups (although not tested by the authors). The study also did not incorporate mtDNA. The authors state that the analysis of concatenated mitochondrial and nuclear DNA "cannot account for coalescences that are deep due to demographic stochasticity rather than to species divergence", though in fact, nuclear loci are more likely to be affected by incomplete lineage sorting since they have a larger effective population size (Hahn, 2018). Nevertheless, both studies chose defensible methods to time-calibrate the macroscelidean tree, yet came up with substantially different evolutionary scenarios.

Although time-calibrating phylogenetic trees with fossils is a common approach in evolutionary biology to reconstruct a taxon's history, differing age estimates among studies are not unusual. To obtain reliable results, calibrating multiple nodes within the phylogeny is desirable (Bibi, 2013). Unfortunately, it is rather typical that the fossil record of a taxonomic group is incomplete, ambiguous or both. Even when fossils are known, their phylogenetic positions are often unclear as a result of the often poor and fragmented condition of many fossil (Sansom et al., 2010), the existence of extinct lineages with unknown phylogenetic positions and missing soft tissue traits used to differentiate recent taxa. Furthermore, uncertainties surrounding the age of a fossil can be another important factor (see below). Due to the incompleteness of the fossil record, fossils only provide minimal ages of clades; in many cases, fossils may be far younger than the true age of clades they are being used to date (Benton and Ayala, 2003; Donoghue and Benton, 2007; Lukoschek et al., 2012; Marshall, 2008). Taken together, when inferring a time-calibrated phylogeny, one faces many pitfalls concerning the credibility of the dating but has little means to circumvent those or even estimate the accuracy of the reconstructed evolutionary scenario. However, it should be mentioned that a more detailed fossil record allows improvement of temporal calibrations through e.g. the fossilized-birth–death model (Heath et al., 2014) as well as approaches exist to deal with uncertainties around placing fossils within the tree (Guindon, 2018). As a result, time-scaled phylogenies are often

published without further evaluation of the biological likelihood of the scenario and without attention to the impact of dataset and parameter settings on the outcome (Graur and Martin, 2004).

Sengis are an example of a clade with a poorly known fossil record. Although a number of extinct crown-group fossils have been described (Holroyd, 2010), none of these could be reliably placed in the sengi phylogeny due to the fragmentary and often highly derived and unique nature of fossil specimens. For this reason, all known sengi fossils can only be used to provide a minimum age for the origin of the entire crown clade (Macroscelididae/Rhychocyonidae, Node 0). Previous studies relied on *Miorhynchocyon meswae* (ibid.) to provide a minimum age of 23–22 mya for dating this node (Heritage et al., 2020; Krásová et al., 2021). More recently described fossil species from Paleogene deposits of Tanzania and Namibia potentially indicate even older ages. *Rukwasengi butleri* from the Nsungwe Formation in the Rukwa Rift Basin of Tanzania is confidently dated to the late Oligocene (25.2 million years (Myr); Stevens et al., 2022). *Namasengi mockeae* from the Eocliff site in the Sperrgebiet of Namibia has been ascribed to the Bartonian-Priabonian (middle to late Eocene, 41.3–33.9 Myr; (Pickford, 2015, 2020; Senut and Pickford, 2021)). This age, however, has been disputed (Marivaux et al., 2014) and tip-dating analyses of Eocliff and other Sperrgebiet rodents have argued for an early Miocene age (22my; (Sallam and Seiffert, 2020)). Unfortunately, Rukwa Rift and Eocliff fossils can again only be used to date the basal Macroscelididae-Rhychocyonidae split (Node 0). Nevertheless, the minimum age for crown sengis can now be assumed to be at least 25.2 Myr and up to 41.3 Myr if Eocliff sengis are indeed of Bartonian-Priabonian age. In this study, we evaluated how different parameters (DNA data set composition, ingroup/outgroup ratio, fossil priors) affect modeled node ages in sengis and whether certain parameter combinations yielded more reliable results. This was done by use of fresh and museum samples that allows us, for the first time, to cover all currently recognized sengi species. DNA sequencing of museum specimens is becoming more common to answer evolutionary biological questions (e.g. Agne et al., 2022b). Methods which are able to deal with the challenges of museum DNA like low endogenous DNA content and its high level of fragmentation have improved substantially over the past years. We cope with these challenges by using laboratory techniques optimized for the recovery of low-quality and low-quantity DNA (e.g. single stranded DNA library preparation (Dabney et al., 2013; Gansauge et al., 2017), target enrichment through in-solution hybridization capture (Horn, 2012; Maricic et al., 2010)) and thereby significantly increase the available genetic data of sengis. If sufficient collection information is available, museum specimens offer the possibility to explore the genetics of populations from remote localities, politically unstable regions or areas

from which the species has been historically extirpated.

4.3 Materials and methods

4.3.1 Dataset

A total of 67 samples were used for this study, 23 museum samples (skins/bones), 6 from fresh tissue, 1 DNA extract from fresh tissue provided by a museum and 37 previously published sequences (Supplementary material, Table S1).

4.3.2 Fresh tissue sample extraction and library preparation

DNA was extracted using a commercial DNeasy kit (QIAGEN, Venlo Netherlands). Extracts were sheared to a target size of 500 bp using a Covaris S220 System (Covaris, Woburn, US-MA) and thereafter converted into double-stranded, double indexed Illumina sequencing libraries (Meyer and Kircher, 2010) (Supplementary material, Table S1).

4.3.3 Museum sample extraction and library preparation

All pre-PCR treatments of museum samples were conducted in a dedicated historical DNA facility at the University of Potsdam. Twenty-three museum samples were processed for this study with collection dates ranging from 1897 to 1979. DNA was extracted following (Dabney et al., 2013) using a non-destructive extraction buffer (Paijmans et al., 2020; Rohland et al., 2004). Extracts from museum samples were built into single-stranded Illumina sequencing libraries (Dabney et al., 2013; Gansauge et al., 2017) with a maximum of 13 ng input DNA.

4.3.4 Target enrichment

Baits for in-solution hybridization capture were built following the protocol of Horn, 2012, Maricic et al., 2010. Template DNA from extracts of three fresh tissue samples was used (*R. petersi* (RHY3717), *E. flavicaudatus* (CAS29703), *M. rupestris* (CAS29696)) to PCR amplify partial coding sequences of nine nuclear loci (A2AB exon 1, ADORA3, ApoB exon 26, ATP7A, BCHE, IRBP, PNOC4, Rag1, vWF). Platinum II Taq Hot-Start DNA Polymerase (ThermoFischer Scientific, Waltham, US-MA) and the manufacturer's recommended protocol were used. PCRs with 45 cycles and primer-dependent annealing temperatures were run (Supplementary material, Table S2). Resulting PCR product sizes were estimated using gel electrophoresis and if they matched the expectations purified using a MinElute PCR Purification Kit (QIAGEN, Venlo Netherlands). DNA concentration and fragment size was determined with a Qubit 2.0 Fluorometer (ThermoFischer Scientific, Waltham, US-MA) and a 2200 Tapestation

System (Agilent Technologies, Santa Clara, US CA). Products of all templates and amplified loci were pooled in equimolar amounts and sheared by sonication to a target size of 150 bp with a Covaris S220 system (Covaris, Woburn, US-MA). Two rounds of hybridization capture with the home-made baits were applied to libraries of all samples except CAS29696, CAS27982, CAS29703, CAS28566, and CAS29752 with an annealing temperature of 65 °C for 24 h following (Horn, 2012; Maricic et al., 2010).

4.3.5   Sequencing

All fresh tissue samples except *E. intufi* (ZFMK0100418693) and *R. petersi* (RHY3717) were shotgun sequenced at a commercial sequencing facility (Novogene, London) to a length of 150 bp paired-end on a NovaSeq6000 system (Illumina, San Diego, US-CA). The latter two samples were processed along with the museum samples after library preparation. All museum samples were shotgun sequenced to approx. 20 M reads 75 bp single-end on an Illumina NextSeq500 system (Illumina, San Diego, US-CA) at the University of Potsdam. Captured libraries were sequenced on the same system to approx. 3M 75 bp single-end reads.

4.3.6   GenBank sequences

To avoid creating chimeric individuals, nuclear sequences of the nine captured nuclear genes of different sengi species were downloaded from GenBank only if voucher specimen information was available. In addition, complete or nearly complete mitochondrial genomes of all available sengi species were downloaded. The raw reads of the available *E. edwardii* genome (SRX8008497) were downloaded from the Sequence Read Archive and treated with the same pipeline as the shotgun sequenced fresh tissue samples after sequencing.

4.3.7   Read processing

Adapters of all raw reads were trimmed using Cutadapt v3.4 (Martin, 2011) and an overlap = 1. Reads <30 bp were discarded. Mitochondrial genomes from samples sequenced on the NovaSeq6000 system were assembled using NOVOPlasty v4.2 (Dierckxsens et al., 2017) with the respectively closest available reference from GenBank (11.16.2020). Mitochondrial sequences of all other samples were assembled using MITObim v1.9.1 (Hahn et al., 2013) running with MIRA v4.0.2 (Chevreux et al., 1999). All assembled mitochondria were annotated with MitoFinder v1.4 (Allio et al., 2020) with *E. edwardii* (NC041486) as reference. To recover nuclear genes,

reads from samples sequenced on the NovaSeq6000 system and of captured libraries were mapped to reference sequences with the program bwa v0.7.17- r1188 (Li and Durbin, 2009; Li and Homer, 2010), using the aln algorithm and reducing the maximally allowed difference to n = 0.01. To circumvent possible mislabeling of museum specimen, reads of each sample were mapped to all available macroscelididae sequences on GenBank for the respective nuclear gene. Duplicates were marked with MarkDupsByStartEnd v0.2.1 (https://github.com/.../MarkDupsByStartEnd) and removed with samtools v1.12 view (Li and Durbin, 2009). Mapping success was compared afterwards and the assembly from the reference with most mapped reads chosen as final result for that locus. Consensus sequences for the nuclear genes were called in two ways. (I) using the samtools v1.12 (Li et al., 2009) mpileup command and (II) more conservatively using Consensify (Barlow et al., 2018). Based on the combined newly-generated plus GenBank sequences, a separate alignment for each gene was produced with MAFFT v7.480 (Katoh and Standley, 2013). Alignments were visually curated, refined and trimmed to the first base of a codon for protein coding sequences in AliView 1.28 (anders, 2014). The resulting gene alignments were concatenated with AMAS (Borowiec, 2016). Partitioning of the data set for each codon position of each gene and the best substitution model for each partition was analyzed with PartitionFinder v2.1.1 (Lanfear et al., 2016) with the following parameters: branchlength = unlinked, models = mrbayes, model_selection = BIC, search = greedy.

4.3.8   Phylogenetic informativeness profiles

To test the data set for substitution saturation, phylogenetic informativeness (PI) profiles for each gene alignment were produced using tapir (https://github.com/faircloth-lab/tapir) (Dornburg et al., 2014; Townsend and Leuenberger, 2011). As reference tree, the outcome of a preliminary BEAST analysis (next section) with *R.butleri* as sengi crown fossil, concatenated mitochondrial and nuclear data and all outgroups was used (see below). If PI profiles peaked before 40 Ma, the top 30% variable positions were masked by replacing the respective columns in the alignment with Ns with a custom-written python script.

4.3.9   Phylogenetic reconstruction

Phylogenetic trees were calculated using maximum likelihood and Bayesian approaches. RaxML v8.2.12 (Stamatakis, 2014) was run with the data partitioning recommended by PartitionFinder (6 partitions) and GTR + G + I as the substitution model for all partitions. 1000 bootstrap replicates were performed. The Bayesian

analysis was conducted with MrBayes v3.2.7 (Ronquist et al., 2012). Convergence could not be reached with the substitution models recommended by PartitionFinder. In order to reduce parameterization and reach convergence HKY + G was used as substitution model for all partitions. Two chains with 100,000,000 MCMC generations, sampling every 10000th step, were run and the first 25% discarded as burn-in after convergence was confirmed with Tracer v1.7.2 (andrew et al., 2018).

4.3.10   Dating

Time-calibrated phylogenies were calculated using BEAST2 (Bouckaert et al., 2014) with two independent, relaxed log normal molecular clocks for mitochondrial and nuclear data. Default settings were used for all clock rate associated starting parameters and estimated by BEAST during the analysis. PartitionFinder recommended six partitions and GTR + I + G for four partitions, GTR + G for one and HKY + G for one, but as BEAST runs did not converge under GTR + I + G, we used HKY + G for all partitions, to reduce parameterization. Lognormal age distributions were chosen for all fossil priors (Ho and Phillips, 2009; Marshall, 2019). Since fossils only provide a minimum age for the origin of a clade, and since clade maximum ages cannot be determined precisely, the standard deviation (M and S parameters) was adjusted such that the 95% age quantile of the prior was 1.25 times the minimum age (see Bibi, 2013; Hempel et al., 2021). Since lognormal priors are 'soft-bounded', estimated ages could easily exceed the 95% quantile we set. For all age priors, the minimum age of the oldest known fossil was used as the offset parameter. To investigate the effect of different fossil priors and data set compositions on the time-calibration, we evaluated the following parameter sets:

- I The fossil calibrating the origin of crown macroscelididae was alternately set to either *M. meswae* (22 Ma; 23.3–27.5 95%) (Holroyd, 2010), *R. butleri* (25.2 Ma; 26.4–31.5 95%) (Stevens et al., 2022) or *N. mockeae* (38.5 Ma; 39.3–48.2 95%) (Senut and Pickford, 2021). Although the age of N. mockeae from Eocliff, Namibia, is disputed (see introduction), it was included as an extreme value to test the effect of different crown macroscelidean fossil priors.

- II The effect of data set composition was investigated by using: 1, all samples (five outgroup and 62 (if mitochondrial DNA was included) or 49 (if only nuclear DNA was used) ingroup samples); 2, no outgroup samples; 3, five ingroup and five outgroup samples; 4, only five ingroup samples. For the latter two approaches, the five representatives of sengi lineages for which maximal data of the target loci were available, were chosen: *E. brachyrhynchus*

(ZMB080087), *E. myurus* (ZMB033146), *G. rufescens* (SMNS28317), *M. micus* (CAS29752), *R. petersi* (RHY3717).

- III The effect of using mitochondrial or nuclear DNA only, or concatenated mitochondrial and nuclear DNA.

All possible combinations of the parameters above were tested, which resulted in 36 trees. In order to simplify the description of the resulting age estimates and their discussion, those 36 dating attempts can be summarized in different dating categories. We tested four different series of dating attempts. Each series differed from the others by its sample composition. Those compositions consisted of either all sengi samples and outgroups (multiple fossil priors), all sengi samples without outgroups (single fossil prior), only five sengi samples and outgroups (multiple fossil priors) or only five sengis samples without outgroups (single fossil prior). If a series included a single fossil prior, it was always the sengi crown-fossil prior (CFP). Each dating series consists of three scenarios, one for each type of DNA tested, either nuclear-only DNA (ncDNA), mitochondrial-only DNA (mtDNA) or concatenated nc and mtDNA (nc + mtDNA). For each scenario three dating attempts were conducted, one for each of the three sengi crown-fossil priors M. meswae, N. mockeae or R. butleri. Taken together a scenario consists of three dating attempts, and three scenarios make up a dating series, hence nine dating attempts.

All dating attempts containing outgroup sequences were calibrated on four additional nodes across the afrotherian tree. Those fossils were (with minimum age and 95% age range in brackets) *Ocepeia daouiensis* (59.2 Ma; 95% 59.7–74.0) for crown Afrotheria (Gheerbrant et al., 2014, 2001), *Eritherium azzouzorum* (Thanetian) (56 Ma; 95%: 56.4–69.9) for crown-Paenungulata (Benton et al., 2015; Gheerbrant, 2009), *Daouitherium rebouli* (55.9 Ma; 95%: 56.4–69.9) for crown-Tethytheria (Kocsis et al., 2014; Seiffert, 2007) and *Todralestes variabilis* (56 Ma; 95%: 56.6–70) for crown-Afroinsectivora (Gheerbrant et al., 1998; Seiffert, 2010). BEAST was run for 90,000,000 MCMC generations, sampling every 10,000th generation. Convergence was checked with Tracer v1.7.2 (andrew et al., 2018). The maximum credibility tree was produced with Tree Annotator v2.6.3 (Drummond and andrew, 2007) after discarding 10–30% MCMC iterations as burn-in.

Rate evaluation

The outcome of multiple dating approaches resulted in large variation of branch-specific substitution rates [substitutions per position per million years] across the tree. To visualize this and detect potential outlier rates, median substitution rates for each terminal in the tree were calculated across each branch from root to tip.

4.4   Results and discussion

Conservative consensus calling with Consensify (Barlow et al., 2018) (data not shown) did not change the phylogenetic topology compared to consensus calling with the samtools mpileup command (Li and Durbin, 2009). All further analyses were therefore conducted with the larger data set acquired via samtools. We extended the previously available genetic data set of sengis with sequences of 29 samples from 18 species and uploaded them to GenBank (Supplementary material Table S1). In total, 20 new complete or nearly complete sengi mitochondrial genomes and 9 partial ones were reconstructed. Our data increase the available nuclear data for all sengi species except *Rhynchocyon udzungwensis* and *Elephantulus pilicaudus*. The final alignment of all nuclear genes, without outgroups, had a length of 8,613 bp of which 1,372 positions were parsimony informative. No saturation of substitutions within the nuclear data set was detected. In contrast, the mitochondrial data set showed clear signs of substitution saturation. The phylogenetic informativeness of most mitochondrial genes peaked before the assumed age of Node 0 (20–40 Ma) (Supplementary material, Fig. S1), which indicates an enhanced probability of the accumulation of hidden substitutions (Dornburg et al., 2014; Townsend and Leuenberger, 2011). After masking the 30% most rapidly evolving positions in the alignment with Ns, PI profiles did not show signs of saturation anymore (Supplementary material, Fig. S1). The resulting mitochondrial alignment of reduced variability and without outgroups had a final length of 14,023 bp of which 2,493 positions were parsimony informative. All analyses were therefore conducted with the masked mitochondrial data set. Due to the different origin and quality of samples and sequencing data (GenBank, fresh tissue samples, museum samples) the amount of total information per sample varied from 21,541 bp for *R. petersi* (RHY3717) to 900 bp for *E. pilicaudus* (EU136167). A median of 13,609 bp + 4,124 bp, 13,561 bp + 5,629 bp and 3,149 bp + 1,786 bp of mitochondrial + nuclear data was retrieved from fresh tissue, museum and GenBank samples, respectively. The final mitochondrial data set consisted of 62 sengi samples, the nuclear one of 49.

4.4.1   Topology

RaxML and MrBayes runs with the concatenated mitochondrial and nuclear DNA data set (nc + mtDNA) resulted in the same tree topology, except for the relationships among subspecies within *Rhynchocyon* (Fig. 3.1). Within soft-furred sengis, all branches but one above species level had very high support values. Within giant sengis, branches were less well supported. Three giant sengi sequences from GenBank clustered with other species than the ones as which they are identified on GenBank

(KT438478, KT348478, MH056212). Nonetheless, the topology of the recently published sengi phylogenies could be confirmed (Heritage et al., 2020; Krásová et al., 2021). Three of our museum samples were identified as mislabeled (MAM31978 is *E. brachyrhynchus* instead of *E. fuscipes*, ZMB033146 is *E.myurus* instead of *E. rupestris*, and ZMB003732 is *E. rupestris* instead of *E. edwardii*.
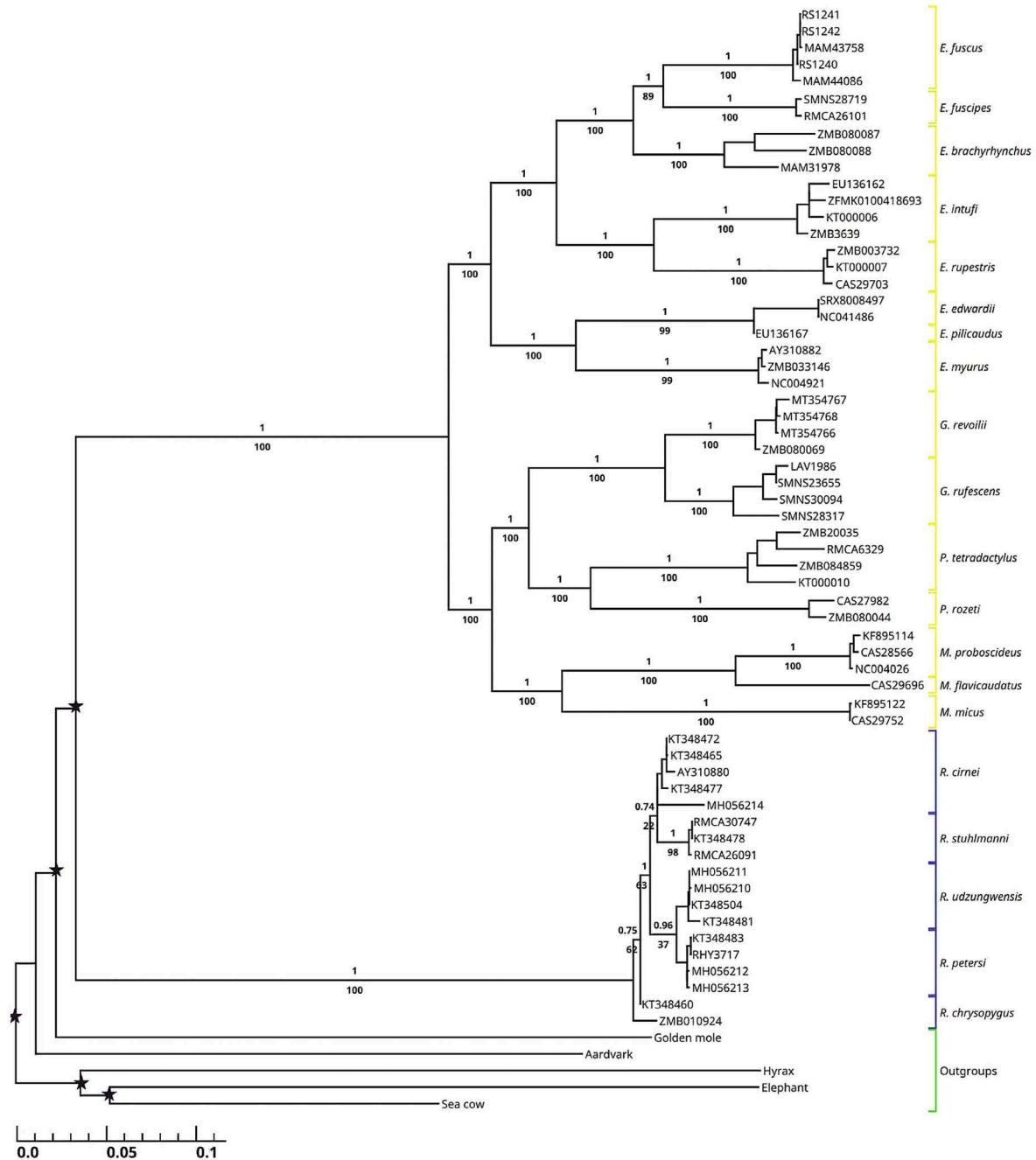
FIGURE 4.2: Phylogeny of Macroscelidea based on nuclear and mito-
chondrial data. Posterior probabilities are indicated above branches,
bootstrap support below. Stars indicate nodes used for age-calibration.
For better visualization, branches outside Macroscelidea are artificially
shortened and do not represent actual genetic distance. Soft-furred
sengis (Macroscelididae) are marked yellow, giant sengis (Rhynchocy-
onidae) blue, outgroups green. (For interpretation of the references to
colour in this figure legend, the reader is referred to the web version of
this article.)

Topologies calculated with nuclear or mitochondrial-only DNA were similarly
well resolved and resembled those produced with the concatenated data set with

minor differences (Supplementary material, Fig. S2 a-d). Previous to this study, the position of *E. fuscipes* was largely unclear as besides a short stretch of mitochondrial DNA, no genetic material was available for this species. We retrieved one complete and one nearly complete mitochondrial genome from museum specimens and up to 4,941 bp of nuclear DNA for our samples of *E. fuscipes*. Relationships among *E. brachyrhynchus*, *E. fuscus* and *E. fuscipes* differed depending on the dataset analyzed (mtDNA/ncDNA/nc + mtDNA). While the sister relationship of *E. fuscus* and *E. fuscipes* had high support in the nc + mtDNA and mtDNA-only datasets, ncDNA data instead suggested a sister relationship of *E. brachyrhynchus* and *E. fuscipes*, with *E. fuscus* being sister to the latter pair. However, all analyses confirmed the monophyly of *Elephantulus* after the recent assignment of *revoilii* and *rufescens* to the new genus *Galegeeska* (Heritage et al., 2020; Krásová et al., 2021). The three *E. brachyrhychus* specimens are deeply divergent, mirroring the geographic distance among sampling localities (ZMB080087 Botswana, ZMB080088 Malawi, MAM31978 DR Congo). *G. rufescens* individuals also separated into two genetic clusters, a northern one from South Sudan and Ethiopia (LAV1968, SMNS23655, SMNS30094) and a southern one (SMNS28317), matching the results of Krasova et al (2021). Similarly, we found deep geographic divergences among the P. tetradactylus individuals. These findings of spatial genetic clustering or the existence of cryptic subspecies/species within these species highlight the importance of further research, especially including museum samples.

4.4.2 Age-calibration

Four out of the 36 dating attempts did not converge (multiple parameters with ESS <200). These used either nc or mtDNA, *R. butleri* or *M. meswae* as sengi crown priors, and came from the analyses that used only five ingroup and five outgroup taxa. Multiple runs of each of these four dating attempts resulted in inconsistent node ages across the repeated runs. They were therefore not considered further. All other dating attempts using either ncDNA or mtDNA converged fully (ESS >200). In the analyses using nc + mtDNA, one or two rate-associated parameters prevented full convergence in three dating attempts (Supplementary material, Table S3). However, multiple runs of these same analyses produced time-trees that were highly congruent, despite not having reached convergence (data not shown). All other dating attempts using nc + mtDNA fully converged. The summarized dating outcomes in the text below refer to the mean age HPD of the respective node. Full node age density distributions are displayed in Fig. 3.2 and supplementary material (Table S3). The topology of phylogenies produced with BEAST resembled those produced with

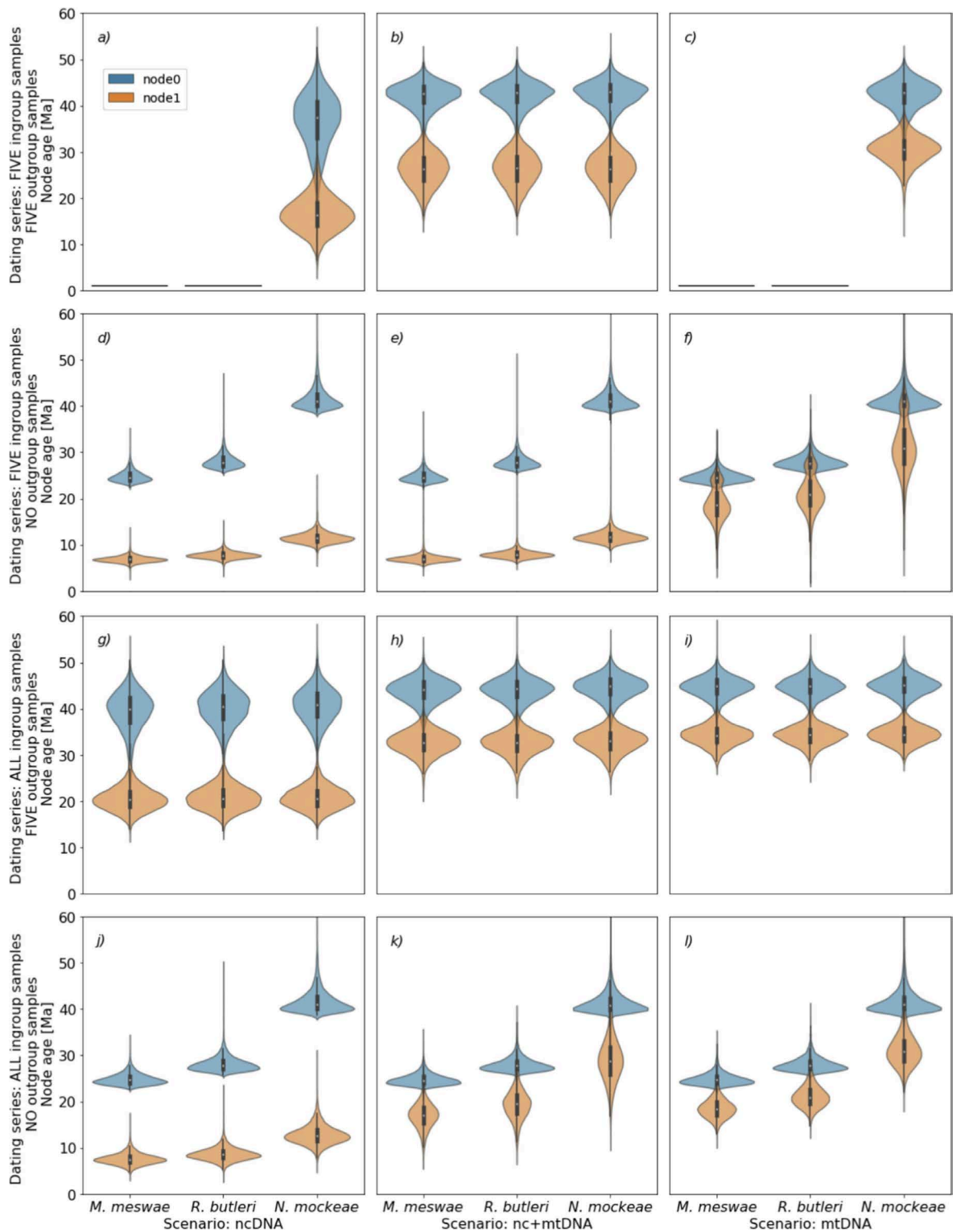RaxML and MrBayes, with minor differences in regard of the position of *petrosaltator* and *fuscipes*.

FIGURE 4.3: Violin plots of age densities (y-axis) of Macroscelidea (Node 0,blue) and Macroscelididae (Node 1, orange) from different BEAST runs. Each line displays a dating series, differing by it's sample composition. Each plot (scenario) covers the three different sengi crown-fossil priors tested in this study (x-axis). Plots in left column: ncDNA, central column: nc + mtDNA, right column: mtDNA. Dating attempts without data in a and c did not converge. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

To test the effect of the age of the CFP on the age of sengis, the impact of three different fossils was assessed. Although the ages of the three fossils used to calibrate node 0 differ by up to 16.5 Myr, an apparent effect on the time calibration could only be observed when no outgroups (and thus no additional fossil priors) were included (Fig. 3.2). Under these conditions, the age of Node 0 reflected the age of the fossil used to calibrate it (Fig. 2. d-f and j-l). The age of Node 1 also followed this trend, i.e., it was comparatively older when using an older calibration than a younger one. If outgroups and thus additional fossil priors were included, the maximal difference, within any single dating scenario, was only 1.8 Myr for Node 0 and 1.2 Myr for Node 1 (Fig. 3.2. a-c and g-i). These findings reveal that, as long as multiple (outgroup) fossil priors are included, the age of the sengi CFP has no major influence on the estimated age of this clade. This is probably because of the large age uncertainties associated with all three fossil priors and highlights the need of new fossils of Macroscelidea, particularly well dated and phylogenetically placed ones from within crown sengi families. In order to test the effect of an incomplete ingroup sampling data set on temporal calibration, we reduced the number of ingroup samples (sengis) to five. We therefore compared the dating outcomes of the series using five ingroup with outgroup samples (Fig. 3.2 a-c) vs. the series using all ingroup samples with outgroup samples (Fig. 2 g-i) and five ingroup samples without outgroup samples (Fig. 3.2 d-f) vs. all ingroup samples without outgroup samples (Fig. 3.2 j-l). A pairwise comparison of dating attempts only differing by the number of ingroup samples resulted in a maximal age difference of 4.1 Myr for Node 0. More generally, the age difference of pairwise comparisons was greater for analyses which included outgroups (0.3–4.1 Myr) compared to those without outgroups (0.-0.1 Myr). Therefore, the number of ingroup samples on the dating of Node 0 was negligible. The effect of the number of ingroup samples on the age of Node 1 was similar, with a maximal difference of 6.4 Myr. These results demonstrate that an incomplete taxon data set is not necessarily inadequate for time-calibration of sengis. However, the three pairwise comparisons between nc + mtDNA, without outgroups and either all samples (Fig. 3.2 k) or the reduced ingroup data set of five sengi samples (Fig. 3.2 e) stood out from this picture. The age differences of Node 1 of those comparisons were between 10 and 16.4 Myr. In order to interpret this outcome, the differences in time-calibration due to different DNA data sets (ncDNA, mtDNA, nc + mtDNA) has to be described first. Pairwise comparisons of dating attempts only differing by the usage of either ncDNA or mtDNA resulted in a consistent picture across all attempts (e.g. Fig. 3.2 a vs. c and j vs. l). If only mtDNA was used, the age of Node 0 was up to 5.7 Myr older compared to using ncDNA. The age difference of Node 1 due to the different type of DNA was even more pronounced. Trees derived from mtDNA resulted in

11–19.9 Myr older estimates for Node 1 compared with trees derived from ncDNA. Ages resulting from the use of nc + mtDNA were similar to those using mtDNA (e.g. Fig. 3.2 h and i) in three out of the four dating-series, whereby the maximal age differences for Node 0 and Node 1 were 2 Myr and 4.2 Myr respectively. Within the dating series of only five ingroup samples without outgroups, however, the age estimates from the scenario of nc + mtDNA resembled those of the ncDNA scenario (Fig. 3.2 d and e). Taken together, if the data set consisted of more than five samples, all dating attempts with nc + mtDNA resulted in similar age estimates as the ones from attempts with mtDNA within a dating series. These similarities can be explained by two overlapping effects. The underlying data of the individual samples and the fossil priors of the outgroups as will be shown below. The median number of basepairs of the samples used in this study was 12,975 bp mtDNA but only 2,154 bp ncDNA. Thus, the more samples were included in a single dating attempt, the higher the proportion of mtDNA in the data set, compared to the proportion of ncDNA (mtDNA/ncDNA ratio = approx. 6). A possible explanation for the tendency of the nc + mtDNA scenarios to follow the age estimates of mtDNA scenarios within the same dating series could be that the phylogenetic signal of the ncDNA was overshadowed by that of mtDNA, if more than five samples were used. In the dating series of five ingroup samples without outgroups, five samples with relatively large amounts of data were chosen, with medians of 13,622 bp mtDNA and 6,007 bp ncDNA (mtDNA/ncDNA ratio = 2.3). However, the samples of the dating series with five ingroup samples and five outgroup samples had medians of 13,806 bp mtDNA and 7,613 bp ncDNA (mtDNA/ncDNA ratio = 1.8), and therefore had proportionally more ncDNA compared to the dating series using only five ingroup samples. Accordingly, if DNA dataset composition was the only effect leading to different age estimates, dating attempts from the latter scenario should come up with similar (young) age estimates as the scenario with ncDNA from this dating series as the mtDNA/ncDNA ratio is even lower in the former. This was not the case (Fig. 3.2 g and h), pointing towards an additional effect by the outgroup fossil priors. To test these assumptions, two new dating scenarios were created, both with concatenated nc + mtDNA. To test the effect of the outgroup priors, the dating series with five ingroup and five outgroup samples (Fig. 3.2. h) was repeated, but without the outgroup fossil priors (only outgroup DNA sequences). To test the effect of the proportion of mtDNA to ncDNA in the data set, the nc + mtDNA scenario from the series of five ingroup samples without outgroups was repeated with a different set of ingroup samples. This time, ingroup samples consisting of less ncDNA information as the previously used five ingroup samples were chosen. Those samples had medians of 13,168 bp mtDNA and 2,154 bp ncDNA (mtDNA/ncDNA

ratio = 6.1). The first additional scenario of five ingroup samples and five outgroup samples but without outgroup fossil priors resulted in age estimates similar to its sister scenario of ncDNA without outgroup fossil priors (Fig. 3.3 a, only dating attempts with *R. butleri* as sengi CFP shown), demonstrating the effect of outgroup fossil priors on the time-calibration, regardless of the mtDNA to ncDNA ratio. As expected the second additional scenario showed, that the ratio of mtDNA to ncDNA per sample influenced the dating outcome. When repeating the nc + mtDNA scenario from the five ingroup - no outgroup dating series (Fig. 3.2 e) with samples consisting of a high mtDNA to ncDNA ratio, the resulting age estimates were now similar to those of the scenario of mtDNA from the same dating series (Fig. 3.3 b, only dating attempts with R. butleri as sengi CFP shown). Summing up, with the two additional dating scenarios, we could show that if the ratio of mtDNA to ncDNA per sample is reduced, the resulting time-calibration approaches the ages of the dating attempts with only ncDNA. Given our results, this would mean that time-calibrations based e.g., on whole genomes will result in age estimates similar to those produced here with ncDNA and are likely more accurate than those derived from mtDNA. The largest difference between scenarios with nc or mtDNA within a dating series were found for the age estimates for Node 1. Dating attempts based on mtDNA resulted in a reduced branch length between Node 0 and Node 1, compared to dating attempts based on ncDNA. This effect was also reflected in the two prior studies (Heritage et al., 2020; Krásová et al., 2021). The former study used both types of DNA, resulting in a branch length between the two nodes of 4.2 Myr. The latter study, in contrast, used only ncDNA, resulting in a branch length of 13.9 Myr. Our findings show that time-calibration attempts with ncDNA are likely to better reflect the actual time-frame of sengi evolution.
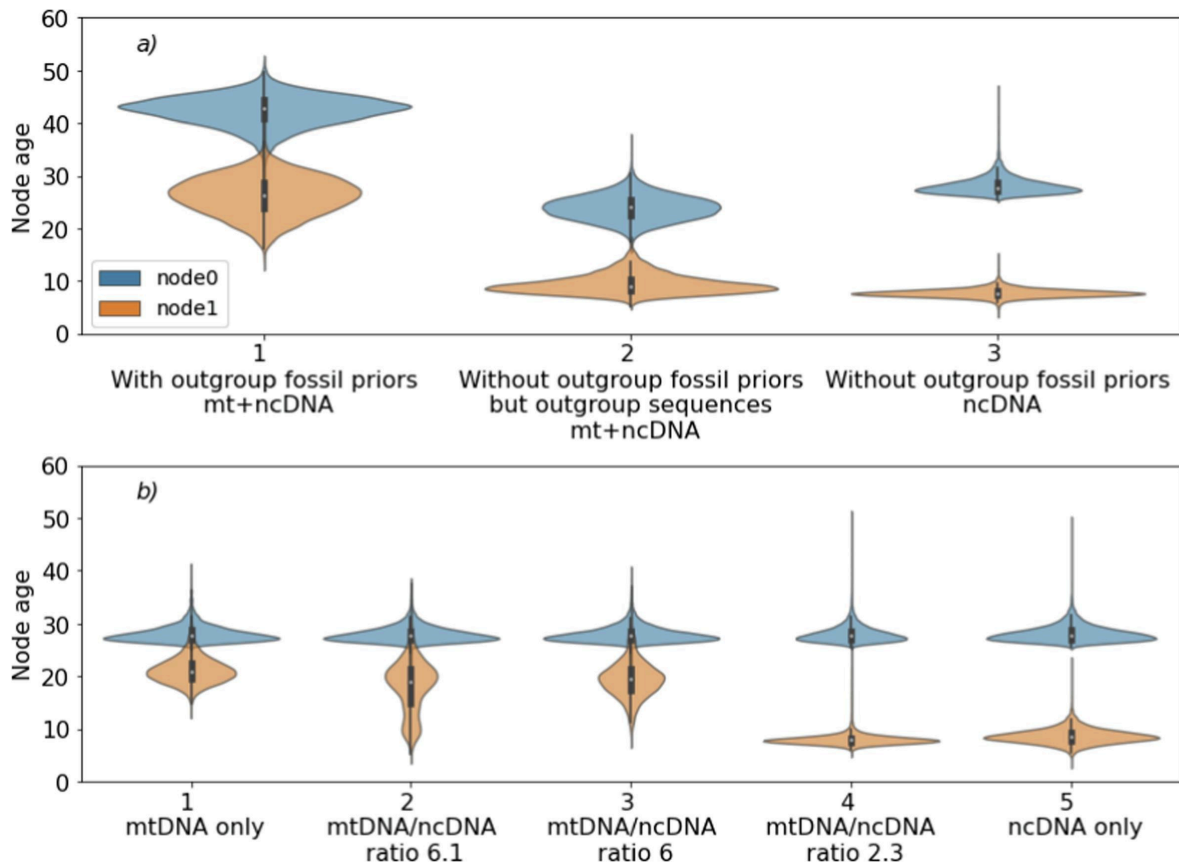
FIGURE 4.4: Violin plots of age densities (y-axis) from different BEAST runs, all dating attempts with R. butleri as sengi crown-fossil prior. a) Effect of outgroup sequences. All three attempts include five sengi samples and if outgroups are included five outgroup samples. 1 and 2 only differ by the inclusion of outgroup fossil priors, but include exactly the same sequences. b) Effect of different mtDNA/ncDNA ratio on node ages. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

However, it is less clear whether the inclusion or exclusion of outgroup fossil priors results in more realistic time-calibrations. In general, it is recommended to recruit multiple fossil priors for temporal calibration in order to calibrate multiple nodes of the phylogeny (Bibi, 2013). In the case of Macroscelidea, these additional nodes are about 20–30 Myr older than the crown clade. In addition, in regard to body size and therefore metabolic rate (Kozłowski et al., 2020), all other afrotherian species except afrosoricidans included in this study are substantially different from sengis, with elephants being the most extreme example. Another means of evaluating the biological likeliness of a dating attempt could accordingly be the substitution rates, gained through the temporal-calibration. In order to test these assumptions, we compared the substitution rates of two dating attempts, both with *R. butleri* as sengi CFP and ncDNA, one with outgroup fossil priors and one without. *R. butleri* was chosen as this is the oldest temporally well-constrained fossil available for

sengis at the moment. The dating attempt without outgroups resulted in an average substitution rate of 0.00379 [substitutions per million years and position]. The same dating attempt including outgroups and thus additional fossil priors, resulted in an average substitution rate of 0.00217 [substitutions per million years and position] for sengis. However, a comparison of terminal specific substitution rates revealed that three branches (Fig. 3.4 a) in the latter phylogeny were associated with either highly elevated or reduced rates. Those outlier rates were 0.0015, 0.0031, 0.0041 on three consecutive branches, followed by a drastic drop back to the average rate without those outlier rates of 0.00185 (Fig. 3.4 a).
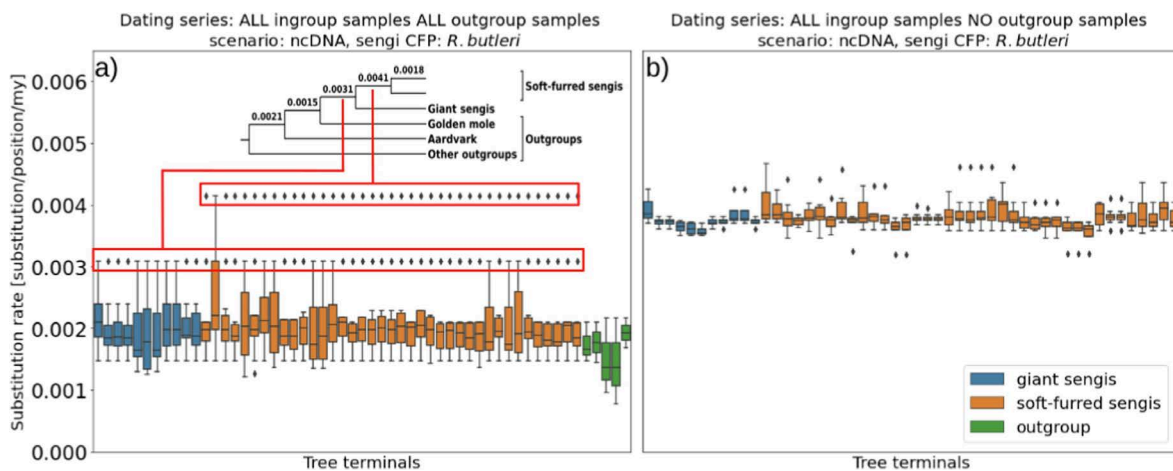
FIGURE 4.5: Boxplot figures: Median branch rates from all branches from root to terminal of the respective tree terminal of the nuclear data set with R. butleri as sengi crown-fossil prior. a) scenario including all ingroup and all outgroup samples (multiple fossil priors). b) scenario without outgroups (single fossil prior). Samples are distributed along the x-axes, median rates on the y-axes in substitutions per million years and position. Simplified cladogram above boxplot graphs in a) shows the corresponding tree with branch specific median substitution rates. Red box and lines indicate which outlier substitution rates (>1.5*IQR) from tree terminals correspond to which branch of the tree. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

We assume that such drastic changes in substitution rate can be considered as biologically unlikely. To put these rate differences into context, the reported difference in mean substitution rate between birds and mammals is, for example, 10% (Kumar and Subramanian, 2002; Nam et al., 2010; Sun et al., 2015). The differences in substitution rates between evolutionary and physiological distant exothermic frogs and endothermic primates are still only approx. 4 fold (Sun et al., 2015). In contrast, the rates from the same dating attempt without outgroups (and associated fossil priors) were more evenly distributed but on average faster (Fig. 3.4 b). Mammalian body size roughly scales with metabolic rate and generation time. Smaller body size

and shorter generation time is associated with faster metabolic rates and substitution rates (Bromham, 2011). However, cross study comparisons of substitution rates are complicated by the inconsistency to estimate and report them (e.g. per generation, different genomic regions) and can vary substantially across the genome (Balmori-de la Puente et al., 2022). In addition, they mostly rely on fossil calibrations and thus on the underlying data set, as shown here. In consequence, reported substitution rates between studies can cover a broad range (Kumar and Subramanian, 2002). More evenly distributed rates across the tree plus the generally faster rate, which would be expected for small mammals (Fig. 3.4 b) favor the dating attempt without additional outgroup fossil priors. We can therefore conclude that the dating attempt with ncDNA, all ingroup samples and *R. butleri* as sengi CFP probably produces the most reliable temporal estimate for sengi evolution for now, dating the age of Node 0 to 28.1 mya (26–31.2 HPD) and Node 1 to 8.7 mya (6–11.5 HPD) (Fig. 3.5, S2).
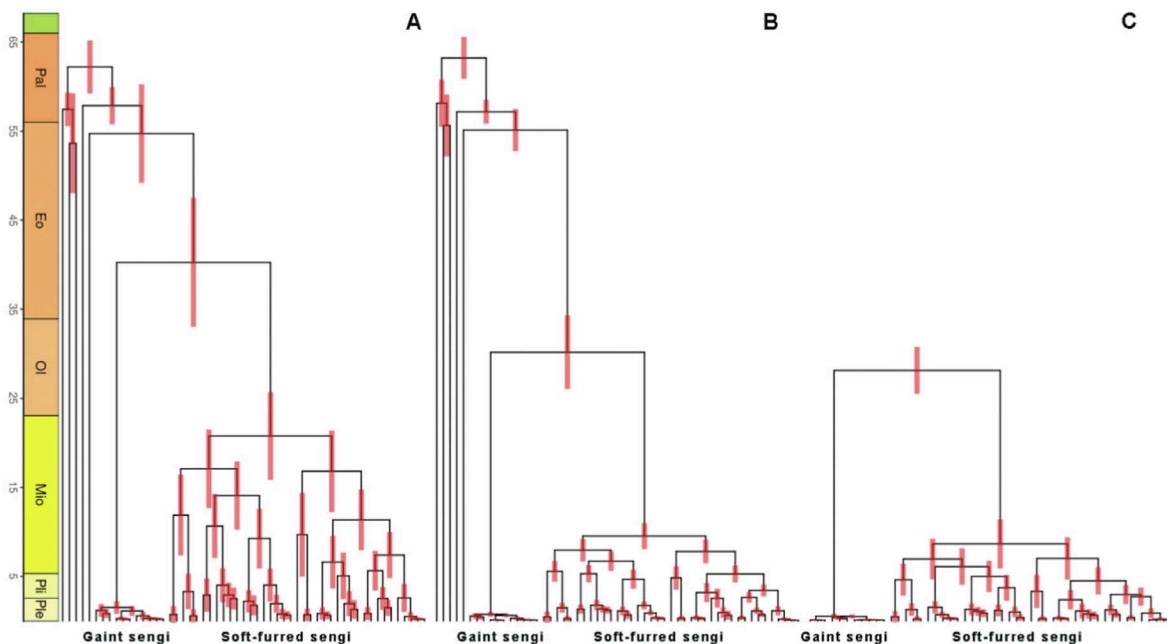
FIGURE 4.6: Effect of molecular clock and dataset composition. All trees are derived from ncDNA and with R. butleri as sengi crown-fossil prior. A) With outgroups (multiple fossil priors) and a log normal molecular clock. B) With outgroups (multiple fossil priors) and a random local clock. C) Without outgroups (single fossil prior) and a log normal molecular clock. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Furthermore, the inconsistent substitution rates across the temporal calibrated trees including outgroup fossil priors, possibly indicate that the chosen clock model (relaxed log normal) is not suitable to model the evolution across such a long time span and biological diverse group. To further investigate this possibility, two additional dating attempts were conducted with ncDNA, *R. butleri* as sengi CFP and

the inclusion of outgroup fossil priors: one employing an optimized relaxed clock (Douglas et al., 2021) and the other employing a random local clock (Drummond and Suchard, 2010).

Using an optimized relaxed clock resulted in similar node age estimates as with a log normal clock (Fig. S2). Using a random local clock, however, resulted in similar young node ages as when not including outgroup fossil priors with a log normal molecular clock. Node 0 was dated to 30.2 mya (26.1–34.4 HPD) and Node 1 to 9.5 mya (8.0–11.0 HPD) (Fig. 3.5, S2). Moreover, the rate estimates were more evenly distributed across the tree, indicating an acceleration towards sengis, which seems biologically reasonable considering the reduced body size compared to e.g. elephants or sea cows. Notably, the uncertainties of the node age estimates were highly reduced in the more shallow nodes by the inclusion of multiple fossil priors (Fig. 3.5). These results further support our conclusion of a rather young evolutionary scenario of sengis. Additionally, they reveal why dating deep nodes within sengis has been notoriously difficult and heavily dependent on data set composition and outgroup set up (see Heritage et al., 2020; Krásová et al., 2021). Overall, we can demonstrate that it is possible and desirable to include multiple fossil priors, even if only distantly related, as long as an appropriate molecular clock model is chosen. A closer examination of the substitution rates and their changes across the tree might be a helpful proxy for this decision.

Our estimates of Node 0 would indicate an Oligocene origin of crown-sengis and therefore strongly suggests that the split between giant and soft-furred sengis was well before the collision of the African and Eurasian continental plate around 23 mya and the subsequent migration of northern placental mammals into Africa (reviewed in Van Couvering and Delson, 2020). Thus, our data (which them-selves are only minimum estimates) show that giant and soft-furred sengi already split within the framework of an isolated, endemic fauna of late Paleogene Africa. Krásová and colleagues (2021) suggested that the repeated opening and closing of savanna corridors associated with the formation of the Rift Valley and a decline in global temperature 6–9 mya served as "speciation pump" for the soft-furred sengi radiation which followed the split of Node 1. Our time estimates of soft-furred sengi radiation basically match with this hypothesis.

4.5   Conclusion

By utilizing museum specimens, we were able to produce a full taxon phylogeny and significantly increase available genetic data of sengis. This study also demonstrated which parameters lead to younger and which to older outcomes when time-calibrating the macroscelidean phylogeny. Besides the highly incomplete fossil record of sengis, we were able to produce a minimal temporal estimate for sengi evolution with high confidence by combining the gained knowledge in combination with a closer investigation of substitution rates. Generally we could show that it is worth the effort to closer examine the data set and its effect on temporal calibration. We recommend that dating studies should generally investigate the effect of the type of DNA, age of fossil priors and inclusion/exclusion of outgroups. Moreover, if biologically very distant organisms are included, a single molecular clock may be unable to estimate biologically-reliable rates and thus a random local clock is possibly the best model choice to include multiple fossil calibration points. Additionally, in order to evaluate the biological likeliness of a temporal calibrated phylogenetic tree, an examination of substitution rates and their change across the tree is recommended.

Our results suggest an Oligocene origin of its two major lineages, soft-furred sengis and giant sengis, and a late Miocene diversification of soft-furred sengis. These results are in accordance with previous dating attempts (Krásová et al., 2021) but younger than others (Heritage et al., 2020). Nonetheless, further fossils assignable to one of the recognized clades within the sengi radiation are needed to improve temporal-calibration for sengis. It also should be confirmed if additional nuclear data on (e.g. on a genomic scale) supports the here produced temporal estimates for sengi evolution, as suggested by our results. Although working on a specific group of mammals, this study demonstrates how commonly varied parameters in temporal calibration of phylogenies affect the outcome of the respective dating and provides means to evaluate the resulting temporal estimates. Dated phylogenies should therefore be seen in the context of the data set which was used to produce them.

4.6   Additional Information

4.6.1   Funding

4.6.2 CRediT authorship contribution statement

Justus Hagemann: Conceptualization, Formal analysis, Investigation, Methodology, Software, Writing – original draft, Data curation, Visualization. **Michael Hofreiter**: Project administration, Supervision, Writing – review & editing, Funding acquisition, Methodology, Conceptualization, Resources. **Faysal Bibi**: Supervision, Writing – review & editing, Methodology, Conceptualization. **Patricia Holroyd**: Writing – review & editing, Conceptualization. **Patrick Arnold**: Project administration, Supervision, Writing – review & editing, Methodology, Conceptualization, Resources.

4.6.3 Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

4.6.4 Acknowledgements

4.6.5 Supplemental Information

All supplemental files of this work are available at:
https://www.sciencedirect.com/science/article/pii/S1055790323000568

Chapter 5

Population genetics of *P. tetradactylus*

Four-toed Sengi (*Petrodromus tetradactylus*, Afrotheria, Mammalia) museomics reveal a crucial role of East African forests in macroscelidean diversification

Justus Hagemann, Luis Victoria Nogales, Michael Hofreiter, Patrick Arnold

5.1 Abstract

Sengis (Macroscelidea) are members of the Afroinsectivora, a group of mammals belonging to the supercohort Afrotheria. Sengis' low population densities and their distribution including politically unstable regions hinder contemporary sampling of comprehensive datasets. We overcome this obstacle for the species *Petrodromus tetradactylus*, one of the most widely distributed sengi species, by utilizing 44 museum samples from multiple natural history museums and creating a dataset of 55 *P. tetradactylus* individuals covering most of the species' distribution. Phylogenetic reconstruction with eleven nuclear loci in conjunction with mostly complete mitochondrial genomes reveals multiple deeply divergent and formerly unknown lineages within this monotypic genus, highlighting the need for a taxonomic revision. Furthermore, we can show that the assumed allopatric distribution of *P. tetradactylus* in central Africa represents most likely a sampling artifact. Biogeographic modeling indicates that the African forest system and its dynamics through climate fluctuations shaped the evolutionary and biogeographic history of this taxon. We show that lineages within *Petrodromus* that were able to adapt to dryer woodland ecosystems are much more widely distributed than lineages restricted to

moist forest systems. Our results suggest multiple parallels between the evolution of the closed-canopy-dwelling soft-furred sengi *P. tetradactylus* and its similar adapted relatives, the giant sengis, like lineages adapted to similar ecotypes with similar distribution.

5.2 Introduction

Sengis (formerly termed elephant-shrews) are a group of small, mostly insectivorous mammals in the Order Macroscelidea Butler, 1956, representing one of the six orders of the clade Afrotheria. All of the currently recognized 20 species are endemic to Africa. Extant sengis can be divided into two families (recently elevated from subfamily level, Senut and Pickford, 2021), Rhynchocyonidae or giant sengis and Macroscelididae or soft-furred sengis. Although the phylogenetic relationships among the living sengi species has been mostly resolved in the recent years (Carlen et al., 2017; Douady et al., 2003; Dumbacher et al., 2016; Hagemann et al., 2023; Heritage et al., 2020; Krásová et al., 2021), little species-specific knowledge of sengis' evolutionary past exists. While the five species of giant sengis are obligate forest-dwellers, the fifteen currently recognized soft-furred sengi species occupy a variety of ecological niches, although they are mostly associated with rocky or sandy, arid and semi-arid habitats with limited vegetation (Perrin and Rathbun, 2013; Rathbun, 2009a). While few soft-furred sengi species also occur to some extent in more vegetated habitats like savanna with dense plant growth, only the Four-toed Sengi, *Petrodromus tetradactylus* Peters, 1846, occupies habitats with closed-canopy woodlands, bushlands and forests. This species does not only differ from other soft-furred sengi species by its closed-canopy-dwelling lifestyle, but also possess morphological traits that clearly differentiate it from other species of the family, such as the name-giving four pedal digits (instead of five), absence of the third (posterior) pair of nipples and also their larger body size, to name some (Corbet and Hanks, 1968; Corbet and Neal, 1965; Heritage, 2018; Jennings and Rathbun, 2001). *P. tetradactylus* has one of the largest distributions among sengis (Jennings and Rathbun, 2001; Rathbun and Dumbacher, 2015). Based on museum records and observations, its distribution can be divided into two allopatric areas. One is covering the central and western Democratic Republic of the Congo (DRC) south of the Congo river and the extreme northern east of Angola, while the second area stretches along the eastern coast of Africa from northern Kenya up to northern South Africa (including the islands of the Zanzibar Archipelago) and inland until Rwanda and Angola (Fig. 1). Four-toed Sengis inhabit forests, dense woody thickets, closed-canopy woodlands, rocky outcrops and riparian areas with precipitation exceeding 700mm (Corbet and Hanks, 1968; Jennings and Rathbun, 2001; Perrin and Rathbun, 2013).
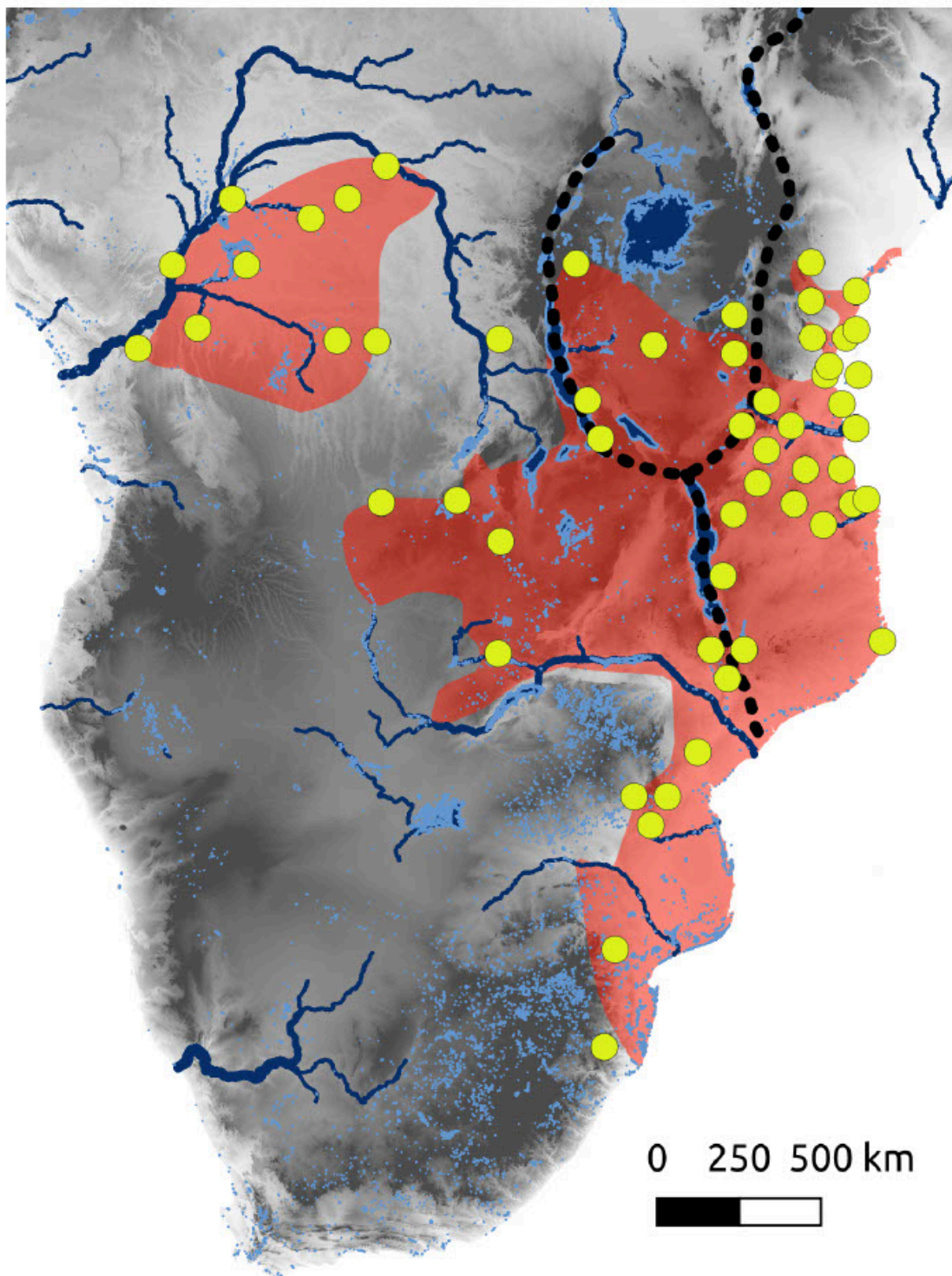
FIGURE 5.1: Topographic map of Sub-Saharan eastern Africa, Red indicating currently accepted *P. tetradactylus* distribution (IUCN Red List, Rathbun, 2015). Blue lines represent rivers and lakes. Yellow circles mark origin of samples used in this study. Black dotted line indicates the East African Rift Valley, whereby the eastern most line represents the course of the Kingdon's Line.

Although monotypic, *Petrodromus* shows some variation in body size, pelage color and tail bristles anatomy across its geographical distribution (Corbet and Neal, 1965). This has historically led to the erection of numerous forms, subspecies and even species (which were later abolished or reduced to subspecies again), with rivers often considered as separating different taxonomic groups (Hollister, 1917; Oldfield, 1897, 1910, 1918). In their extensive examination, Corbet and Neal (1965) reviewed these morphological traits based on sampling locations and geographic distributions of previously proposed forms. They grouped the samples into eighteen areas for comparison. However, as the authors noted, due to limited sample availability, the delimitation of those areas was "rather arbitrary" and did not follow a convincing grouping by specific ecological regions, nor by geographic barriers (Corbet and Neal, 1965). Nonetheless, they found a morphologically distinct form in the coastal forests of southern Kenya and north-eastern Tanzania. They suggest that this coastal form may have been more widespread before the late Pleistocene and was separated from an inland form by a dry corridor through the Luangwa Valley to the southern extension of the East African Rift System in Tanzania. The authors further hypothesized that the inland form later expanded its range and invaded the coastal areas where it interbred with the coastal population, creating a number of different forms along the coast of eastern Africa. The existence of a morphologically more-or-less well characterized forms distributed along the Tanzanian and Mozambique coast, plus the existence of specimens with intermediate phenotypes located in-between the latter form and the northern coastal one are indicative of this hybridization scenario (ibid.). In addition, the authors proposed another evolutionary distinct lineage associated with Congolese lowland forests due to its allopatric distribution in central DRC (Figure 1) and sufficiently sharp morphological differences (Corbet and Hanks, 1968; Corbet and Neal, 1965). Its distinctiveness is generally well accepted in sengi literature (Jennings and Rathbun, 2001; Rathbun, 2009a). The morphological variation and current evolutionary scenarios thus suggest an important role of forests (eastern coastal forest, Congolese lowland forest) in *Petrodromus* diversification. The relevance of the African Sub-Saharan forest system on the evolution of mammals (including humans) has been highlighted by multiple studies in the recent past (e.g., Bryja et al., 2017; Joordens et al., 2019). The main drivers in this system are climatic, geological and ecological events that often interact. While a single forest system stretched across equatorial Africa during the humid early to middle Miocene, this ecosystem was split up by tectonic uplift and the creation of the Rift Valley during the late Miocene (Plana, 2004). Climatic changes in the late Miocene may have caused widespread aridification, resulting in the replacement of forests along the Rift Valley through a savanna-type ecosystem and thus presumably creating a barrier for forest-dwelling

species between the Guinean-Congolese forest in western and central Africa and the forest system east of the Rift Valley. The role of this barrier, which stretches from Ethiopia to Mozambique, has been shown to be so crucial for the evolutionary history of many floral and faunal species, that it, in functional resemblance to the Wallace-Line, was named Kingdon's Line (Fig. 1) (Grubb et al., 1999; Joordens et al., 2019). Climatic fluctuations during the Pleistocene further structured the two main African forest systems east and west of the Kindon's line (Bryja et al., 2017; Joordens et al., 2019; Kingdon, 2013). Glaciations during ice age cycles in the northern hemisphere were accompanied by cooler and dryer climate in Sub-Saharan Africa. The two forest systems were fragmented into island-like patches, separated by savanna-type ecosystems, potentially limiting geneflow between the forest islands for forest dwelling species. During interglacials in the north, the more humid climate around the equator led to expansion of the forest islands, reconnecting them and enabling the dispersal of forest-dwelling species. Throughout the multiple climatic cycles of the Pleistocene, a number of forest refugia persisted, such as the East African coastal forest and the Eastern Arc Mountain forests, offering a more stable ecosystem and allowing the survival of obligate forest dwelling species (e.g., Bryja et al., 2017; Joordens et al., 2019; Kingdon, 2013). DNA sequence analysis provides a tool to investigate the phylogeographic structure of *P. tetradactylus* and test the existence of different evolutionary lineages, their interrelationships and their potential association with forest refugia in eastern Africa. In conjunction with biogeographic modeling, this data allows us to test the plausibility of hypotheses concerning the evolutionary past of *P. tetradactylus* as formulated by Corbet and Neal (1965) based on morphological characters. However, detailed examination of intra-specific diversity and evolutionary history in sengis using molecular data are currently hampered by the limited availability of samples of sufficient quality and quantity. Due to their very low population densities for mammals of their size (approx. 1-2 individuals/ha) (Oxenham and Perrin, 2009) and relatively broad distribution, including politically unstable regions (Fig. 1), the collection of sufficiently large sets of fresh samples for sengi species such as *P. tetradactylus* is difficult. However, natural history museums throughout the world harbor large collections of specimens collected over the past centuries. Recent developments in laboratory techniques enable us to make use of the genetic information within those collections by extracting and analyzing the often fragmented and poorly preserved DNA from tiny tissue samples, while preserving the valuable specimens for future research (e.g. Agne et al., 2022a; De Abreu-Jr et al., 2020; Hofreiter, 2012; Rohland and Hofreiter, 2007; Rohland et al., 2004; Straube et al., 2021). The present study makes use of those collections by processing a total of 44 museum specimens of the soft-furred sengi *P. tetradactylus* in conjunction with

genetic information from public data bases and fresh tissue samples, spanning the species' range (Fig. 1). The extensive dataset thus enables us to investigate the phylogeographic past of the genus *Petrodromus*. We test if multiple divergent lineages exist within this genus and if they are in accordance with previously formulated spatial population structure based on morphological evidence.

5.3 Materials and methods

5.3.1 Samples

Small tissue samples from 44 museum skins stored in a number of natural history museums were processed during this study (known sampling years 1897-1955), as well as a single fresh tissue sample from 1995 (MO32862) (Supplemental Table S1). Genetic information on 10 additional specimens was retrieved from GenBank along with their sampling localities. Thus, 55 *P. tetradactylus* specimens in total were included in this study, resulting in one of the largest datasets for any sengi species available to date. If GPS coordinates of the specimens' origin were not available, they were approximated in Google Maps (Google LLC) from the sampling location indicated on the specimens' voucher (Supplemental Table S1). The samples are distributed across the entire known distribution range (IUCN Red List, Rathbun, 2015) of *P. tetradactylus* (Fig. 1). .

5.3.2 DNA extraction and library preparation

All pre-PCR laboratory processing of museum samples was conducted in a dedicated historical DNA facility at the University of Potsdam. DNA of museum samples was built into single-stranded libraries, see Supplemental text S1 for details.

5.3.3 Bait preparation and hybridization capture

Capture baits for in-solution hybridization capture were constructed for eleven nuclear loci (Adra2B exon 1, ADORA3, ApoB exon 26, ATP7A, BCHE, CREM, IRBP, PLCB4, PNOC4, Rag1, vWF), see Supplemental text S1 for details. Two rounds of hybridization capture were applied to all samples with an annealing temperature of 65°C for 24h following (Horn, 2012; Maricic et al., 2010).

5.3.4 Sequencing and read processing

In order to reconstruct mitochondrial genomes, DNA libraries were shotgun sequenced to approx. 12M 75bp single-end reads on an Illumina NextSeq500 system (Illumina, San Diego, US-CA). Hybridization enriched DNA libraries were sequenced to approx. 3M 75bp single-end reads on the same system. The hybridization enriched library of the only fresh tissue sample used in this study (M032862) was sequenced to approx. 15M 150bp paired-end reads on the same sequencing system. For details on how read processing, mapping, assembly and annotation was conducted, please see Supplemental text S1.

5.3.5   Phylogenetic reconstruction

Alignments of either nuclear-only, mitochondrial-only or concatenated nuclear and mitochondrial DNA were partitioned by gene and codon position with Partition-Finder v2.1.1 (Lanfear et al., 2016). Maximum Likelihood (ML) and Bayesian phylogenetic trees were reconstructed for each dataset with RaxML v.8.2.12 (Stamatakis, 2014) and MrBayes v3.2.7 (Ronquist et al., 2012) and the partitioning recommended by PartitionFinder. For details see Supplemental text S1 for details.

5.3.6   Haplotype Networks

The genetic relationships among samples was additionally tested by the reconstruction of haplotype networks using the Minimum Spanning method (Paradis, 2018) in PopArt (Leigh and Bryant, 2015), see Supplemental text S1 for details.

5.3.7   Temporal calibration

The most comprehensive dataset of concatenated mitochondrial and nuclear DNA did not converge, as observed in other sengi studies (Hagemann et al., 2023), and could therefore not be used for temporal calibration. Unfortunately, the nuclear-only DNA dataset did not contain enough variable positions to unambiguously resolve the phylogeny. Therefore, a dataset of mitochondrial-only DNA had to be used to infer temporal estimates. However, it has been demonstrated that time-calibrated trees from mitochondrial DNA tend to overestimate node ages in sengis (ibid.). To avoid this overestimation and make use of temporal information in the nuclear DNA dataset, we therefore applied a two-step dating procedure.

5.3.8   Primary dating

Nuclear DNA of eleven genes of four *P. tetradactylus* individuals representing the major lineages found in the phylogenetic reconstruction, of all other soft-furred sengi species (except *E. pilicaudus* for which only limited genetic data are available) and of three giant sengi species as outgroup were used for primary dating (Supplemental Table S2). Partitioning of the dataset was done with PartitionFinder v2.1.1, as described above. Beast2 (Bouckaert et al., 2014) was run with a relaxed molecular clock, the Yule model as population tree prior and each partition under the HKY+G model. The Macroscelidea crown node was calibrated with Oligorhynchocyon songwensis (25.2 My; 26.4-31.5 95%) (Stevens et al., 2022) and a log-normal distribution (Offset=25.2my, standard deviation M=1 S=0.51). Chains were run for 50,000,000

iterations. A maximum credibility tree was produced after discarding the first 10%
MCMC iterations and confirmation of convergence, as described above.

5.3.9   Secondary dating

Thirteen mitochondrial protein coding genes and the two rRNA genes of 53 *P.
tetradactylus* individuals were used for secondary temporal calibration. Beast2 was
run as above, except that the crown *P. tetradactylus* node was calibrated with the
age-range extracted from the primary dating approach (1.18-2.11mya), a log-normal
distribution with 0.65my offset, standard deviation M=1.3 and S=0.25.

5.3.10   Biogeography

In order to trace the geographic origin of *Petrodromus'* diversification, its biogeo-
graphic history was analyzed using the R package BioGeoBEARS (Landis et al.,
2013; Matzke, 2014; R core team, 2022). Ancestral ranges at internal nodes were
estimated by allowing a combination of different historical events (vicariance, an-
agenetic dispersal, jump dispersal) among *Petrodromus* lineages. Three different
models of range evolution on a phylogeny were compared with Maximum Like-
lihood: Dispersal-Extinction-Cladogenesis (DECLIKE; based on Ree et al., 2005),
Dispersal-Vicariance Analyses (DIVALIKE; based on Ronquist, 1997), and Bayesian
inference of historical biogeography for discrete areas (BAYAREALIKE; based on
Landis et al., 2013). Each model was implemented with or without jump dispersal as
an additional cladogenetic event (+J parameter; Matzke, 2014). The biogeographic
analyses were performed using the time-calibrated mitochondrial tree reduced to
the major monophyletic lineages within *Petrodromus* (Fig. 2A). Three biogeographic
delimitations were employed according to georegions, biomes and ecoregions in
the Afrotropic. In the first analysis, georegions were delimited according to ma-
jor geographic barriers (rivers, lakes, mountain ranges, Rift valley) that have been
suggested to potentially limit the range of *Petrodromus* populations (Corbet and
Neal, 1965) or eastern African mammals in general (Grubb et al., 1999). Georegions
were defined as 1) Zanzibar archipelago; 2) Kenyan-Tanzanian coastal zone (east of
Kingdon's line, north of Rovuma river); 3) northern Mozambique (south of Rovuma
river, east of Lake Malawi, north of Zambezi river); 4) south of Zambezi river; 5)
east African plateau (west of Kingdon's line, east of Albertine Rift Mountains and
Lake Tanganyika, north of Tanzanian Southern Highlands); 6) Central-Zambezian
(west of lake Malawi and lake Tanganyika, north of Zambezi river, east of upper
Congo/Lualaba river); or 7) south (and west) of Congo river. In the second analy-
sis, *Petrodromus'* range was delimited by large-scale afrotropical biomes according

to Dinerstein et al., 2017 [https://ecoregions.appspot.com/]. Major lineages were coded as either occupying 1) tropical and subtropical grasslands, savannahs and shrublands or 2) tropical and subtropical moist broadleaf forests (and forest mosaics). In the third analysis, a finer-scale delimitation by afrotropical ecoregions based on the same source as above was applied. Major lineages were accordingly coded as either occupying 1) Miombo-Mopane woodlands (including Zambezian-Limpopo mixed woodlands, central Zambezian wet Miombo woodlands and Limpopo lowveld); 2) Swahili coastal forests and woodlands (including Maputaland coastal forests and woodlands); 3) Eastern Arc forests; 4) forest-savannah mosaics (including Congolese forest-savannah and Victoria basin forest-savannah); or 5) Congolese lowland forests. The fits of alternative biogeographic models (DECLIKE, DIVALIKE, BAYAREALIKE) were compared using AICc values and subsequently derived AIC weights for each of the three delimitation schemes (georegions, biomes, ecoregions). We also performed likelihood-ratio tests (LRT) for nested models with and without jump dispersal (+J) for each biogeographic model to assess if adding this parameter improved model fitting.

5.4　Results

5.4.1　Dataset

Laboratory processing of the 45 *P. tetradactylus* individuals yielded a median of 13,365 bp mitochondrial coding DNA sequence per sample and a median of 5,386 bp of nuclear DNA sequence per sample across the eleven loci. No correlation between sample age and genetic information extracted could be observed, but samples from the Royal Museum of Central Africa yielded on average much more information than samples from, for instance, the Museum of Natural History, Berlin or the Museum of Comparative Zoology at Harvard University, which might be associated with the museum's or collector specific preservation methods (not shown). Within the mitochondrial dataset, 2,800 out of 13,960 positions (about 20%) were parsimony informative but only 188 out of 7,707 positions (about 2.4%) in the nuclear dataset, highlighting the importance of faster evolving mitochondrial sequences for characterizing within-species phylogeny if nuclear data is limited. Due to the low variability within the nuclear dataset, phylogenetic analysis using nuclear-only DNA could only reliably resolve the deeper splits within the phylogeny (Supplemental Fig. S1). The result and discussion section will therefore focus on topologies derived from the concatenated nuclear and mitochondrial dataset, if not stated differently.
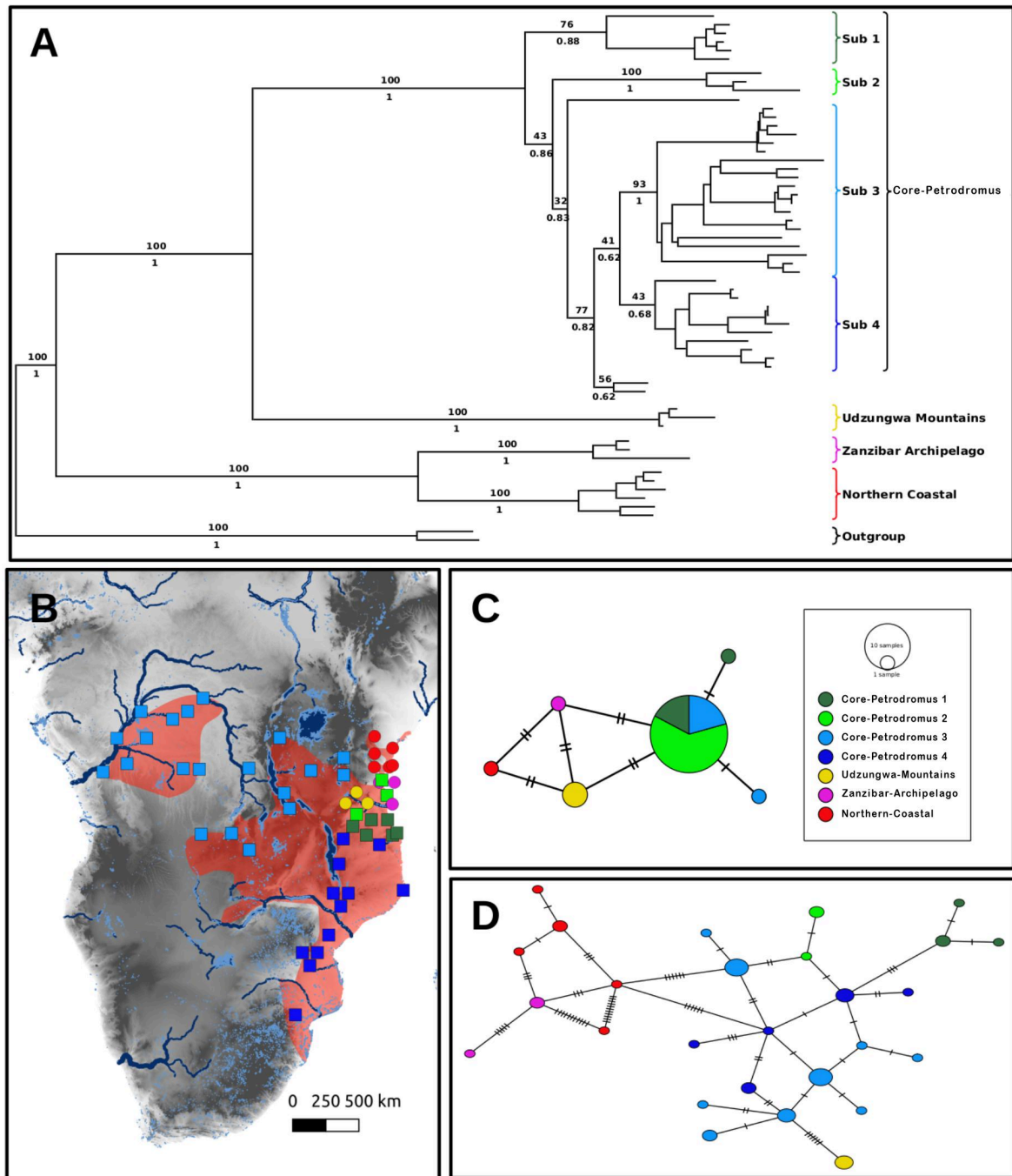
FIGURE 5.2: Phylogenetic structure of *P. tetradactylus*. Color coding indicates the main phylogenetic lineages found. **A**: Phylogenetic tree from mitochondrial and nuclear DNA with major genetic lineages highlighted. Numbers above branches denote bootstrap support, below branches posterior probabilities. Three samples with limited amounts of data were not assigned to a specific lineage. **B**: sampling locations of samples assigned to lineages, squares represent the Core-*Petrodromus* lineage, dots all other lineages. **C**: Haplotype network of the nuclear gene IRBP. **D**: Haplotype network of the mitochondrial gene 16S-rRNA.

5.4.2   Phylogenetic reconstruction

ML and Bayesian phylogenetic reconstruction resulted in congruent phylogenetic trees. Four major phylogenetic lineages within the genus *Petrodromus* were uncovered (Fig. 2A). One lineage consists of samples from the Zanzibar Archipelago (from here on referred to as Zanzibar Archipelago lineage), one lineage unites individuals from coastal southern Kenya to northeastern Tanzania (Northern Coastal lineage), a third lineage consists of individuals sampled in the Udzungwa Mountains National Park (Udzungwa Mountains lineage), while the fourth lineage is constituted by individuals sampled from basically the whole range of *P. tetradactylus* (Fig. 2). During this study we address the latter lineage as Core-*Petrodromus* lineage, as the type locality (Tete, Mozambique) of *P. tetradactylus* falls within this clade (Peters, 1846). All topologies (nuclear and mitochondrial DNA) suggest a basal split between the Core-*Petrodromus* lineage and the Zanzibar Archipelago/Northern Coastal lineage clade. Whether the latter two lineages are monophyletic or whether one is nested within the other could be not fully resolved by the nuclear-only dataset (Supplemental Fig. S1). The position of the Udzungwa Mountains lineage could not be thoroughly determined. Topologies reconstructed from mitochondrial-only DNA and concatenated nuclear and mitochondrial DNA highly supported a sister relationship between the Udzungwa Mountains lineage and the Core-*Petrodromus* lineage (1 posterior probability (PP), 100% bootstrap support (BSS)). However, topologies reconstructed from nuclear-only DNA and all *P. tetradactylus* samples placed the Udzungwa Mountains lineage basal to the Zanzibar Archipelago/Northern Coastal lineage clade with medium support (0.91 PP, 80% BSS). The Core-*Petrodromus* lineage can be further split-up into four sub-lineages (Core-*Petrodromus* sub-lineage 1-4, Figure 2). Due to limited data, the position of three samples (HS85 (South Africa), RS1204 (Zambia), M5X106 (South Malawi)) within the Core-*Petrodromus* lineage could not be reliably resolved and are therefore not assigned to a specific lineage in Figure 2A and not displayed in panels B, C and D in Figure 2.

5.4.3   Haplotype networks

The phylogenetic structure found above was further supported by haplotype networks. Within the network of the 1582 bp long 16S rRNA gene alignment with 414 variable sites, the Udzungwa Mountains lineage is separated from the Core-*Petrodromus* lineage by six mutations (Fig. 2D). The Northern Coastal lineage together with the Zanzibar Archipelago lineage are separated from the Core-*Petrodromus* lineage by six mutations and the former two from each other by at least three mutations. The Core-*Petrodromus* sub-lineages 1 and 2 from eastern and southern Tanzania

are respectively separated from all other lineages by at least one mutation, but the widely-distributed sub-lineages 3 and 4 cannot be resolved with this short DNA fragment. In the network derived from the 954 bp nuclear IRBP gene alignment all the four main lineages are separated from each other by at least two mutations, but the Core-*Petrodromus* sub-lineages are not clearly separated from each other (Fig. 2C). The haplotype network of full mitochondrial genomes showed a similar picture, the one of vWF was not variable enough to separate the Nothern Coastal lineage, the Zanzibar-Archipelago lineage and the Core-*Petrodromus* sub-lineage 1 from each other (Supplemental Fig. S2).
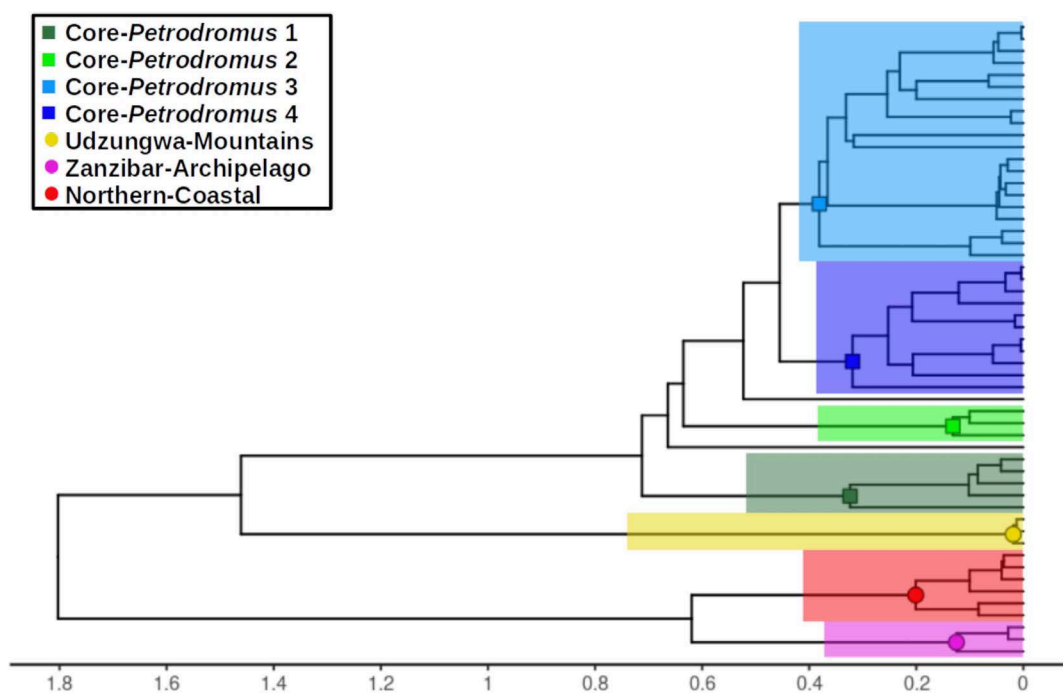
FIGURE 5.3: Time-calibrated phylogenetic tree estimated from mito-chondrial DNA. Major lineages and color coding follow Figure 2, x-Axis in million years ago. Lineages marked with squares show the Core-*Petrodromus* lineage, Yellow dot Udzungwa lineage, Red dot Northern Coastal lineage, Purple dot Zanzibar Archipelago lineage. The samples (HS85 and RS1042) which could not be assigned to a lineage in the topology part due to little data are not color coded.

5.4.4   Evolutionary Time Frame

Primary dating with nuclear-only DNA estimated the split between *P. tetradactylus* and its sister species *Petrosaltator rozeti* to be 4.99 mya (95% CI 3.9-6.23 mya). The crown-node of *P. tetradactylus* was estimated to be 1.63 mya (95% CI 1.18-2.11 mya)

(Supplemental Fig. S1). Notably, within this dataset the sister relationship between the Udzungwa Mountains and Core-*Petrodromus* lineages had maximal support (1 PP). Subsequent secondary dating with mitochondrial-only DNA estimated the split between the Zanzibar Archipelago lineage and Northern Coastal lineage to 0.62 mya (95% CI 0.34-0.93 mya) and the one between the Udzungwa Mountains lineage and the Core-*Petrodromus* lineage to 1.46 mya (95% CI 0.83-2.12 mya) (Fig. 3). The four sub-lineages within the Core-*Petrodromus* clade diverged between 0.71 and 0.46 mya from each other (Fig. 3). The seven genetic lineages uncovered here and in the phylogeny section were used for biogeographic analysis. The position of the two samples HS85 (South Africa) and RS1042 (Zambia) could not reliably be placed within any of the here inferred lineages by BEAST and are therefore not color coded in Figure 3 as well as excluded from the biogeographic inferences below.

5.4.5 Biogeography

Likelihood-ratio tests (LRT) revealed that the addition of jump dispersal (+J) significantly improved the DIVALIKE model in the biome analysis and the BAYAREALIKE model in the georegion analysis (Supplemental Table S3). AIC weights nevertheless suggest that the BAYAREALIKE (without J) model is most likely for the biome and ecoregion analysis and the DIVALIKE model (without J) for the georegion analysis. Ancestral range estimation for the georegion analysis revealed that all extant *Petrodromus* sub-lineages had their origin most likely and with high probability in the Kenyan-Tanzanian coastal zone, where all but one of them still exist today. A vicariance event is suggested in the ancestor of the Northern Coastal and Zanzibar Archipelago lineages that subsequently divided them into a coastal and an island form. The range of the ancestor of the Core-*Petrodromus* sub-lineages 3 and 4 could not be estimated reliably as the Kenyan-Tanzanian coastal zone alone or in combination with the east African plateau and/or the Central-Zambezian georegion were assigned with similar but low probabilities. However, at least two dispersal events starting from the Kenyan-Tanzanian coastal zone were estimated to have taken place in the middle Pleistocene since the divergence of Core-*Petrodromus* sub-lineages 3 and 4 from the Core-*Petrodromus* sub-lineage 2: on the one hand, an eastward dispersal to the east African plateau, the Central-Zambezian and to south (and west) of the Congo river in sub-lineage 3; and on the other hand, a southward dispersal into northern Mozambique and to the region south of the Zambezi river (as far south as northern South Africa) in sub-lineage 4 (Fig. 4).
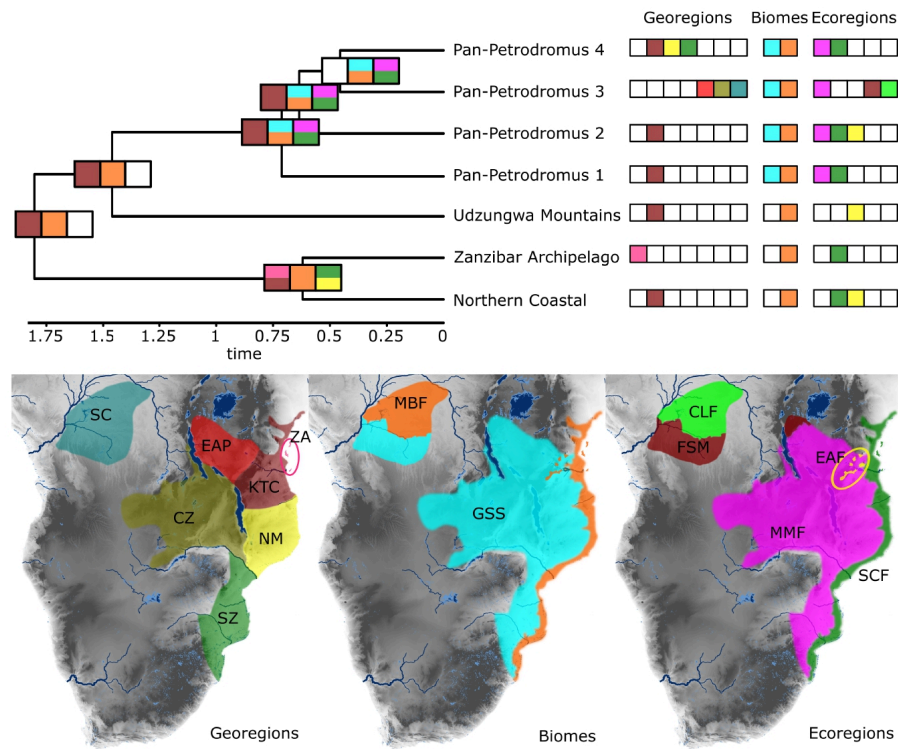
FIGURE 5.4: **Georegions: CZ**: Central-Zambezian, **EAP**: East African Plateau, **KTC**: Kenyan-Tanzanian coastal zone, **NM**: Northern Mozambique, **SC**: South of Congo river, **SZ**: South of Zambezi river, **ZA**: Zanzibar archipelago. Biomes: **GSS**: Tropical and subtropical grasslands, savannahs and shrublands **MBF**: Tropical and subtropical moist broadleaf forests. **Ecoregions: CLF**: Congolian lowland forest, **EAF**: Eastern Arc Forests, **FSM**: Forest-savannah mosaics, **MMF**: Miombo-Mopane woodlands, **SCF**: Swahili coastal forests and woodlands. Squares at tree nodes visualize estimate of ancestral ranges for, from left to right, the georegion, biome and ecoregion analyses, respectively. White squares indicate that no reliable ancestral range could be estimated.

Range estimates for the biome analysis suggested a moist broadleaf, forest-dwelling origin for the ancestor of living *Petrodromus* lineages with high probability. The same biome was estimated for the ancestor of the Northern Coastal and Zanzibar Archipelago lineages as well as the ancestor of Udzungwa Mountains plus Core-*Petrodromus* sub-lineages. The analysis hypothesized an early to middle Pleistocene dispersal event into the grassland, savannah and shrubland biome in the ancestor of all Core-*Petrodromus* sub-lineages, which all range in both biomes (Fig. 4). Ancestral range in the finer-scaled ecoregion analysis could not reliably be estimated either for the ancestor of all living *Petrodromus* lineages or the ancestor of Udzungwa Mountains plus Core-*Petrodromus* lineages as the model could not discriminate between different combinations of Miombo-Mopane woodlands, Swahili coastal forests and Eastern

Arc forests. The ancestor of the Northern Coastal and Zanzibar Archipelago lineages was estimated as restricted to Swahili coastal forests and Eastern Arc forests and the ancestor of the four Core-*Petrodromus* sub-lineages as restricted to Miombo-Mopane woodlands and Swahili coastal forests. The ecoregion analysis suggested two dispersal events: the first into Eastern Arc forests (secondarily) in the Core-*Petrodromus* sub-lineage 2; and the second into Forest-Savannah mosaics and Congolese lowland forests in the Core-*Petrodromus* sub-lineage 3 during the middle Pleistocene (with accompanying retraction from Swahili coastal forests and woodlands in the latter).

5.5   Discussion

We were able to produce a comprehensive genetic dataset from museum samples of the genus *Petrodromus*, spanning its entire known distribution. We could again highlight that museomics allows to gather datasets for studying the molecular evolution of species for which the acquisition of modern samples in sufficient quantity is difficult. As observed in other studies, the genetic information a sample yields does not depend on its age, but more likely on the preservation method used and therefore often the museum where it is housed (e.g. Straube et al., 2021). For sengis, the Royal Museum of Central Africa yielded the best preserved samples for the extraction of genetic material.

5.5.1   Phylogenetic structure

We found multiple deeply diverged lineages within this taxon, some of which correspond to morphological variation described by Corbet and Neal (1965) while others fail to do so. In their intensive study of morphological variation within *Petrodromus*, Corbet and Neal (1965) found two distinct lineages associated with the Kenyan/northern Tanzanian coast and the Zanzibar archipelago. Both lineages have been granted subspecies status (*P. t. sultani* and *P. t. zanzibaricus*; Corbet and Neal, 1965) and the former one even (temporarily) species status (*P. sultani*; Thomas 1897). These morphological results are congruent with our molecular ones as we also identified two divergent lineages at the northern coast and on the Zanzibar Archipelago, respectively. The exact taxonomic status of these two lineages thus is in need of taxonomic revision in the future, taking both lines of evidence into account. Contrary to further findings by Corbet and Neal (1965), we found no evidence that specimens from central DRC south of the Congo river represent an evolutionary distinct lineage within *Petrodromus* – despite their deviating morphology and isolated distribution that led previous authors to the erection of subspecies (*P. t. tordayi*; ibid.) or even

species status (*P. tordayi*; Thomas 1919). As demonstrated in Figure 2, spatial proximity within *Petrodromus* does not necessarily indicate genetic proximity. Individuals sampled close to the East African coast (e.g. Core-*Petrodromus* sub-lineages 1 and 2, Fig. 2), are genetically closer to individuals sampled in DRC (Core-*Petrodromus* sub-lineage 3) than to the spatially closer individuals from the Northern Coastal lineage. We thus confirm Dumbacher and colleagues (2016), who first suggested that Tanzanian specimen are not necessarily closely related (but based on four individuals only). Furthermore, our data shows that no obvious genetic separation between individuals from DRC and western East Africa (i.e., west of the Eastern Rift valley/Kingdon's line) exists, as individuals from both assumed allopatric distributions cluster in the Core-*Petrodromus* sub-lineage 3 (Fig. 2). To our knowledge no apparent ecological differences or geographic barriers exist which would explain the assumed absence of *P. tetradactylus* in south-eastern DRC. The supposed discontinuous distribution of *P. tetradactylus* might therefore more likely reflect a historical sampling artifact than actually allopatric populations (as already suggested by Corbet and Neal (1965)). This explanation is supported by the sampling location of RMCA23458 in Bosumba, DRC which lies in between the two areas (Fig. 2). However, future sampling efforts should verify this claim. The molecular results further support the hypothesis of a very recent dispersal (<40ka) into the Congo basin (Jennings and Rathbun, 2001). We also found another deeply diverging *Petrodromus* lineage within the Udzungwa Mountains National Park that has so far not been recognized based on morphology (Fig. 2 and 3). This is particularly interesting, as a recently described new giant sengi species, *Rhynchocyon udzungwensis*, occurs in the very same location (Rovero et al., 2008). Our results are further evidence of the unique and rich biodiversity within this area (Davenport et al., 2006; Dinesen et al., 2001; Kunene et al., 2022; Rovero and De Luca, 2007). These findings together with the fact that all the four main lineages described here exist in close spatial proximity, or even sympatry in the Tanzanian Coastal zone, highlight the need for comprehensive taxonomic revision of the genus *Petrodromus*. Morphological comparisons of these lineages are necessary to evaluate their taxonomic and consequently also their conservation status (which is beyond the scope of the presented work).

5.5.2 Biographic history and evolutionary timeframe

In contrast to previous assumptions (e.g., Jennings and Rathbun, 2001), major geographic barriers like rivers, lakes and mountain ranges, do not fully explain the genetic structure of *Petrodromus* (Fig. 2). Instead, the biogeographic analyses suggest a crucial role of forests in early *Petrodromus* diversification, particularly in the

Kenyan-Tanzanian coastal zone east of the Eastern Rift valley/Kingdon's line. In recent years, increased attention has been payed to the role of climatic change and its effect on forest distribution and connectivity across Africa during past geological epochs (Bryja et al., 2017; Couvreur et al., 2008; Joordens et al., 2019; Krásová et al., 2021; Trauth et al., 2005). During the Miocene (approx. 23-5mya) central Sub-Sahara Africa was mostly covered with forests. Influenced by glaciations of the northern hemisphere during the late Miocene, the climate in Sub-Saharan Africa got dryer and cooler, leading to a fragmentation of the formerly connected forests and thus creating a more island like habitat distribution for forest dwelling species (Bobe, 2006; Bryja et al., 2017; Joordens et al., 2019; Plana, 2004). Over the multiple glacial cycles during the Pleistocene, those forest islands grew and shrunk, getting connected and separated (e.g. Potts, 2013). However, multiple stable forest refugia existed throughout this period, like the east African coastal to Eastern Arc mountain forests. These forest refugia allowed forest-dwelling species to exist throughout the dryer climatic periods, while wetter periods allowed dispersal across the reconnected forests (Bryja et al., 2017; Joordens et al., 2019).
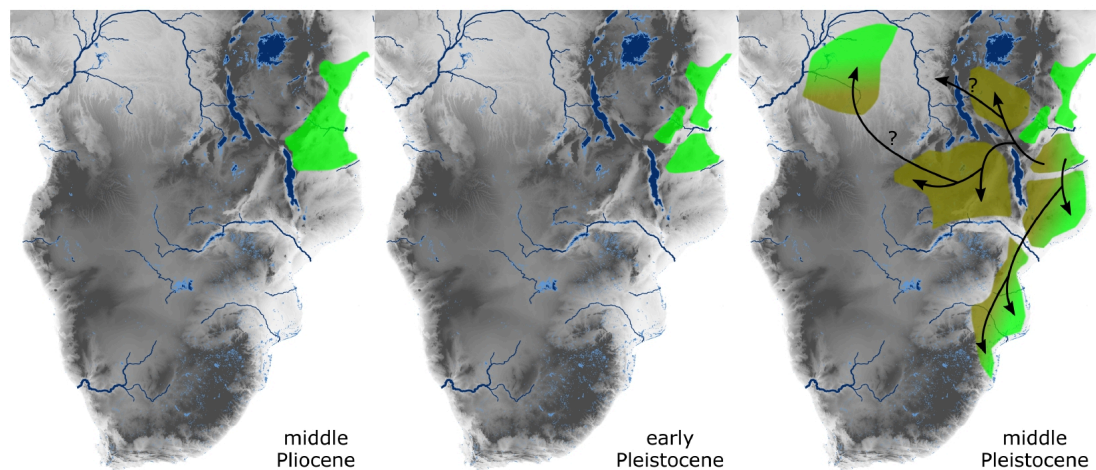
FIGURE 5.5: Biogeographic past of the genus *Petrodromus* throughout past geological epochs

Our biogeographic analysis indicates that the evolution of *Petrodromus* was shaped

by those dynamics, too. Although molecular dating attempts have to be interpreted carefully, the split between *P. tetradactylus* and its sister species *P. rozeti* most likely dates to around 5 mya (3.9-6.2 95% CI), which falls within the period of late Miocene aridity. Similar ages were estimated by other studies (Hagemann et al., 2023; Krásová et al., 2021). This time period has been shown to be of importance for the evolution of other mammals inhabiting the east African coastal forest (Bryja et al., 2017). Increased aridity and the opening of arid corridors in eastern and southern Africa has been suggested to have shaped speciation processes in soft-furred sengis before (Hagemann et al., 2023; Krásová et al., 2021). As *P. rozeti* is a dryland adapted sengi, our results indicate that this adaptation to dryer ecosystems resulted in the separation of the two species during a time when major bioregions formed in Africa (Linder et al., 2012). *P. tetradactylus* survived arid climate phases in a humid forest refugium east of the Kingdon's line, most likely in the East African coastal forest, while *P. rozeti* adapted to dryer ecosystems. During the Pilocene, the climate was more humid again (Couvreur et al., 2021; Kingdon, 2013), reconnecting the previously fragmented forest east of the Kingdon's line, which remained a dryland corridor and thus ecological barrier (Fig. 1) (Grubb et al., 1999; Joordens et al., 2019). Studies analyzing pollen in drilling cores indicated dry climate periods from 1.9-1.7mya and 1.1-0.9 mya (Morley and Kingdon, 2013; Trauth et al., 2005). The former fits with our dating of the crown-node of *Petrodromus* (1.82 mya), suggesting that forest fragmentation might have initialized early *Petrodromus* diversification. We therefore suggest the following hypothesis on the species' evolutionary diversification: *Petrodromus* originated in the East African coastal forests (Fig. 5). Due to the buffering effect of the Indian Ocean, the East African coastal forest was less affected by dryer climate and is regarded as a stable habitat throughout the climatic changes of the Pleistocene (Joordens et al., 2019), facilitating its refugial role for endemic species (Burgess et al., 1998). During wetter climate periods in the Pliocene, the stem-*Petrodromus* population was able to expand inland up to Kingdon's line through the forests that were reconnected then, probably along the riverine forests. Dry periods during the middle Pleistocene fragmented the forest system east of the Kingdon's line, which lead to a split of the stem-*Petrodromus* into a northern and southern lineage, possibly along the Rufiji river (Northern Coastal vs. Udzungwa Mountains plus Core-*Petrodromus* sub-lineages, respectively; Fig. 4) . The common ancestor of this northern lineage inhabited the coastal forest as well as northern Udzungwa Mountains and was then separated by a vicariance event into a coastal mainland and a Zanzibar archipelago lineage. The ancestor of the southern lineage populated the forests of southern Eastern Arc Mountains (including Udzunga Mountains) and coastal forest (probably south of

the Rufiji river). From the latter, the ancestor of the Core-*Petrodromus* lineage expanded into dryer ecosystems like savanna and woodland biotopes. This shift to dryer and less vegetated ecosystems (e.g., Miombo woodlands) facilitated a southern and western dispersal into similarly dryer habitats, but nevertheless did not hinder a colonization of the moister rain forest of the Congo basin (Fig. 4 and 5). This proposed evolutionary history has some parallels to that of the second major lineage of sengis, the obligate forest-dwelling giant sengis (*Rhynchocyon*). Within this taxon, species are closely related but each species is associated with a different east African forest (Amin et al., 2021; Carlen et al., 2017; Lawson et al., 2021). Three species have limited distributions in Kenyan and northeastern Tanzanian coastal forests (*R. chrysopygus*, *R. petersi*) and the Udzungwa Mountains (*R. udzungwensis*). Two other species are more widely distributed: *R. cirnei* inhabits coastal forests of southern Tanzania and northern Mozambique as well as forests and woodlands east and west of lakes Tanganyika, Rukwa and Malawi (Fig. 6). It is therefore the only giant sengi species which occours in dryer closed-canopy woodland (Perrin and Rathbun, 2013). The last species, *R. stuhlmanni*, in contrast, occupies lowland rainforests in northeastern DRC. Although the radiation of *Rhynchocyon* is considered to be more recent than the one of *Petrodromus*, parallel biogeographic lineages have formed in both groups to some extent (Fig 6). This is most obvious for *R. udzungwensis* and the here uncovered Udzungwa Mountains lineage in *Petrodromus*. The distribution of *R. petersi* is also very similar to the combined Northern Coastal/Zanzibar Archipelago lineages, particularly as *R. petersi* is divided in a mainland and island subspecies (*R. p. petersi* and *R. p. adersi*, respectively). Core-*Petrodromus* sub-lineage 4 is reminiscent of *R. cirnei* in extending southward to the Zambezi river. Similarly, with Core-*Petrodromus* sub-lineage 4 and *R. stuhlmanni* there is a single lineage in each taxon that colonized Congolian lowland forests. *R. stuhlmanni* is closely related to *R. cirnei* and has only recently been raised from a subspecies of the latter (Carlen et al., 2017). Together, *R. cirnei/stuhlmanni* with its close relationship but wide geographic distribution (from Tanzanian coast to the south and into Congo forests) is quite similar to the Core-*Petrodromus* lineage in general (but has been more constrained to dense woodlands and forests). *Petrodromus* additionally shows high genetic diversity in south-eastern Tanzania (Core-*Petrodromus* sub-lineages 1 and 2), which is not mirrored in *Rhynchocyon*. Our data do not reveal whether there is also a separate *Petrodromus* lineage in Kenyan northern coastal forests that would be equivalent to *R. chrysopygus*, as no samples from this region were included. However, both species occur sympatrically in the Kenyan Arabuko-Sokoke forest (Fitzgibbon, 1995), which calls for sample collection there in the future (Fig 6). Altogether, both *Petrodromus* and *Rhynchocyon* share high phylogenetic diversity in the Eastern African coastal region

east of the Eastern Rift valley/Kingdon's line (Lawson et al., 2021, 2013; Rovero et al., 2008) while at the same time in each group, only a single nested lineage with wide geographic distribution exists south of the Rovuma river (and in *Petrodromus* even south of the Zambezi river) and west of the Rift valley extending as far as into Congolese lowland forests. Interestingly, the morphological peculiarities that distinguish *Petrodromus* form other soft-furred sengis are also shared with Rhynchocyon, e.g., larger body size and reduction of phalange I.
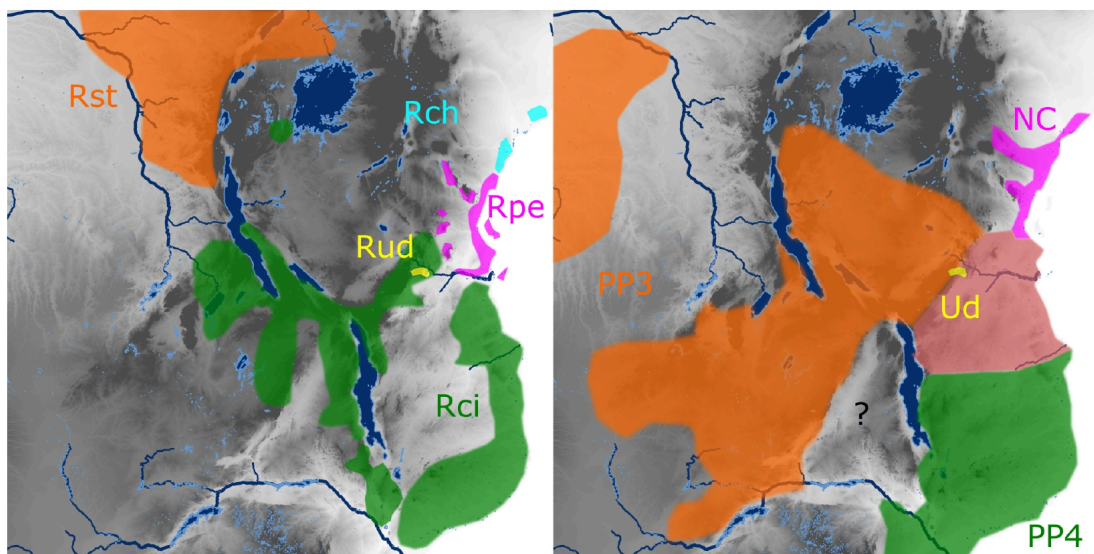
FIGURE 5.6: Biogeographic parallelism in forest-dwelling sengis, *Rhynchocyon* (left) and *Petrodromus* (right). *Rhynchocyon* species: Rch *R. chrysopygus*, Rci *R. cirnei*, Rpe *R. petersi*, Rst *R. stuhlmanni*, Rud *R. udzungwensis*. *Petrodromus* lineages: NC Northern coastal, PP3/4 Core-*Petrodromus* sub-lineage 3/4, Ud Udzungwa Mountains lineage

Although not fully contemporary, the ecological and biogeographic parallelisms between *Rhynchocyon* and *Petrodromus* suggest that Pleistocene cycles of expansions/retractions of forest refugia in eastern Africa played a crucial role in lineage diversification in sengis. This is surprising as sengis are known for being particularly diverse in arid areas (apporx. 75% of all species, with specialized boulder, gravel plain or sand dwellers in different regions; Rathbun, 2009a). The late Miocene opening of arid corridors in eastern and southern Africa has therefore been suggested as the pivotal event in extant (soft-furred) sengi diversification ("speciation pump";

Krásová et al., 2021). Our molecular results for *Petrodromus* in combination with previous findings for *Rhynchocyon* (Amin et al., 2021; Carlen et al., 2017; Lawson et al., 2021), however, revealed that the forest refugia, which formed during Miocene aridification and later expanded/retracted cyclically during the Pleistocene, similarly fueled lineage diversification in two closed-canopy forest and woodland-dwelling clades among sengis.

5.6 Conclusion

By utilizing museum specimens for phylogenetic analysis, we were able to assemble a comparative dataset allowing the reconstruction of the biogeographic and evolutionary past of the genus *Petrodromus*. We could show that expansions and contractions of the East African forest system shaped the evolution of *Petrodromus* during Pleistocene climatic fluctuations and not geological barriers like rivers, lakes and mountains, as previously assumed. Though likely not contemporary, *Petrodromus* evolutionary history shows some parallelism with the similarly forest-dwelling giant sengis of the genus *Rhynchocyon* in their phylogeographic patterns, which again highlights the so far neglected role of closed-canopy forests and woodlands in sengi diversification. Several deeply divergent genetic clades plus an up to now unrecognized lineage from the Udzungwa Mountains implies the need for taxonomic revision of the genus *Petrodromus* by combining morphological and molecular data.

5.7 Additional Information

5.7.1 Funding

5.7.2 CRediT authorship contribution statement

Justus Hagemann (Performed research, wrote first draft, analyzed data, designed research), Luis Victoria Nogales (Performed research, analyzed data, wrote paper), Michael Hofreiter (Funds acquisition, wrote paper) and Patrick Arnold (Designed research, performed research, funds acquisition, wrote paper).

5.7.3 Declaration of Competing Interest

All authors declare no conflict of competing interest.

5.7.4 Acknowledgements

5.7.5   Data Accessibility statement

All sequences produced for this study are publicly accessible. Please see Supplemental Table S1 for NCBI GenBank accession numbers and sequences.

5.8 Supplements

Please refer to the electronic version of this work to view the supplemental files, as they are in no printable format.

Chapter 6

Comparative genomics

A genomic approach on the enigmatic evolution of sengis (Macroscelidea)

Justus Hagemann, Michael Hofreiter, Rohan Nath, Arunkumar Krishnan, Stefanie Hartmann, Patrick Arnold

6.1 Abstract

Sengis (Macroscelidea) represent one out of six extant orders of the Afrotheria, a major mammalian clade with an evolutionary history closely entangled with the African continent. Our knowledge and understanding of the evolutionary past of sengis is rather incomplete, and multiple aspects about sengis appear enigmatic or contradictory in the light of evolutionary theory.

Scientific insights on sengis so far are mostly based on morphological and behavioral studies. Genetic studies included only a few loci and thus provided only enough information to answer basic taxonomic questions. Whole-genome sequencing and comparative genomics offer a new approach to investigate sengi evolution and might help to detect aspects about sengi evolution overlooked by more traditional approaches.

In this study, we generate whole-genome sequencing data for thirteen out of the twenty sengi species and analyze this data set in combination with another previously published sengi genome and multiple ones from other afrotherian clades. By using TOGA, a new method for the annotation of mammalian genomes, we were able to produce a data set consisting of more than 8000 protein-coding one-to-one orthologs present in every sengi genome for phylogenomic analysis. Besides investigating the demographic history of sengis in comparison to those of other small African mammals, we also take a closer look at the evolution of the olfactory sense, a trait regarded to be of importance for sengis. By identifying different genes which, exclusively for a single phylogenetic clade, either experienced positive selection, gene-family expansion, gene-family contraction, or were lost while contributing to the same mammalian phenotype, this study takes a novel approach to detect phenotypic traits which might be of adaptive value for different phylogenetic clades within sengis.

The here gathered comprehensive genomic dataset allowed us to produce a high-confidence temporal calibration for sengis, with considerably reduced node age uncertainties. In contrast to many studies on other mammalian clades, we did only detect a negligible amount of gene tree discordance and thus no sign for extensive past-speciation gene-flow. Furthermore, we identify multiple candidate genes and associated phenotypic traits which might be of adaptive value for different phylogenetic groups within sengis. The physiological nature of these phenotypes indicates that species-specific differences might be cryptic and therefore overlooked by previous research.

6.2   Introduction

Sengis (Macroscelidea, formerly termed elephant shrews) are an order of Afrotherian mammals currently consisting of 20 extant species. Although not particularly poorly studied in the past (Rathbun and Woodall, 2002), sengis show such a puzzling combination of traits, some seemingly conflicting in the light of evolution, that our understanding of their evolutionary history is far from being satisfying.

Sengis consist of two well-defined families, the giant-sengis (Rhynchocyonidae, five species) and the soft-furred sengis (Macroscelididae, fifteen species). They are characterized by a unique combination of life-history traits like small size and low body weight < 1kg, extended noses and long tongues, large eyes, mouse-like tails, highly cursorial locomotion, exposed sheltering habits (in soft-furred sengis), a predominantly myrmecophage diet, no reported ability to climb, e.g., on trees, small litter sizes (one to three), mostly two precocial (in soft-furred sengis) offspring, and minimal parental care for neonates usually consisting of a single maternal visit for lactation per day (Rathbun, 1979, 2009b; Rathbun and Rathbun, 2006). Some of these traits are otherwise found in such disparate species as small antelopes, like Dik-Diks on the one hand and ant-eaters on the other hand.

The evolution of sengis is restricted to the African continent and the origin of this clade is assumed to be of considerable age (Holroyd, 2010). Their MRCA is currently dated to around 30mya and their split from the next closest Afrotherian clade, the Afrosoricida (golden moles, otter shrews, and Malagasy tenrecs), was dated to about 56mya (Gheerbrant et al., 1998; Hagemann et al., 2023; Seiffert, 2010). The fossil record is limited but revealed multiple extinct lineages, with some fossils as old as the early Eocene (Holroyd, 2010; Senut and Pickford, 2021; Stevens et al., 2022). However, it documents very little morphological change over the past 30-40mya, as the overall cranial and postcranial morphology of sengis has not changed much since at least the earliest Miocene (Novacek, 1984; Pickford and Senut, 2008; Senut and Pickford, 2021). Only the transition from a herbivorous diet to an insectivorous one, testified by extant sengis still having a working caecum, is well supported by a change in tooth morphology (Rathbun, 2009b; Woodall and Mackie, 1987).

As Africa was isolated from other continents until approx. 21mya before the formation of the Tauride land bridge (Pekar and Deconto, 2006; Van Couvering and Delson, 2020), sengis are assumed to have been evolving without or with little competition for similar niches from other small mammals like rodents (Rathbun, 2009b), which potentially could explain the limited morphological evolution testified by the fossil record. However, after the connection of Africa and Eurasia, one would either expect a dispersal of sengis out of Africa or an invasion of similarly adapted mammals

from Eurasia into Africa, leading to enhanced competition for the same ecological niches and some kind of evolutionary effect, leaving traces in sengis' phylogeny, morphology, or both. No indication exists that sengis ever dispersed out of Africa, but there have been multiple emigrations of Eurasian small mammals into Africa (Van Couvering and Delson, 2020). The fossil record documents substantial sengi diversity in the Early Miocene (23-18mya), most of which is lost afterwards (Butler, 1995; Pickford and Senut, 2008; Senut and Pickford, 2021). From around 13mya only three sengi species are known (see Holroyd, 2010). Whether this loss in diversity is connected to climatic changes, Afro-Laurasian faunal exchange, or both is not clear. Nonetheless, the two recent main lineages of giant sengis and soft-furred sengis survived without substantial morphological changes within them. Furthermore, the main radiations within extant Macroscelidea substantially postdate (soft-furred sengi radiation 8-1mya and giant-sengi radiation 0.5mya) major waves of Laurasian immigration (Van Couvering and Delson, 2020) and are more likely connected to climatic variations (Bryja et al., 2017; Hagemann et al., 2023; Lawson et al., 2013), suggesting that sengis represent two evolutionary successful lineages in the context of competition with other small mammals for similar niches.

However, although the 20 extant sengi species (especially the 15 soft-furred sengis) occupy a broad variety of habitats, ranging from rocky outcrops, desert gravel plains and arid woodland to mesic savannas and tropical rainforest, no morphological traits exist which would explain the differences in habitat use between the species within the two extent families. Moreover, within soft-furred sengis, this species-specific restriction to different habitats also does not follow any obvious phylogenetic pattern. For instance, the various rupicolous (rock dwelling) species are not necessarily phylogenetically closer to each other than to species living in other habitats (Heritage, 2018; Perrin and Rathbun, 2013). This fact could indicate that sengis are evolutionarily flexible enough to adapt to widely different habitats without large morphological change. On the other hand, if this were true, it should be expected that they would have dispersed out of Africa upon the connection between Africa and Eurasia, similar to Eurasian mammals extending their range into Africa. In this context, the fact that no sengi species has been recorded in sub-Saharan western Africa is even more puzzling (Corbet and Hanks, 1968; Heritage, 2018; Perrin and Rathbun, 2013).

Taken together our knowledge of sengis is currently insufficient to reconstruct and understand their evolution. However, research focusing on sengis has been mostly based on studies of morphology and behavior (e.g. Andanje et al., 2010; Rathbun, 1979; Rathbun and Rathbun, 2006). Genetic studies so far focused on phylogenetic aspects and are lacking a sufficient amount of data to draw any conclusions regarding functional aspects of sengi biology or adaptation (Carlen et al., 2017; Douady et al.,

2003; Hagemann et al., 2023; Heritage et al., 2020; Krásová et al., 2021; Smit et al., 2011). Genomics may allow a better understanding of Macroscelidean evolution, especially with respect to traits that are too cryptic to be detected on a morphological or behavioral level. In light of the low level of their morphological diversity, the taxonomic diversity within sengis with different species living in habitats with widely differing conditions regarding parameters like climate, substrate, and the amount of vegetation could potentially have its basis in physiological differences invisible on the morphological or behavioral level. Since it is not immediately obvious, which physiological parameters and underlying genetics could be important, a comparative genomics approach could present a fresh opportunity to gain insights into the enigmatic evolutionary history of this unique animal group.
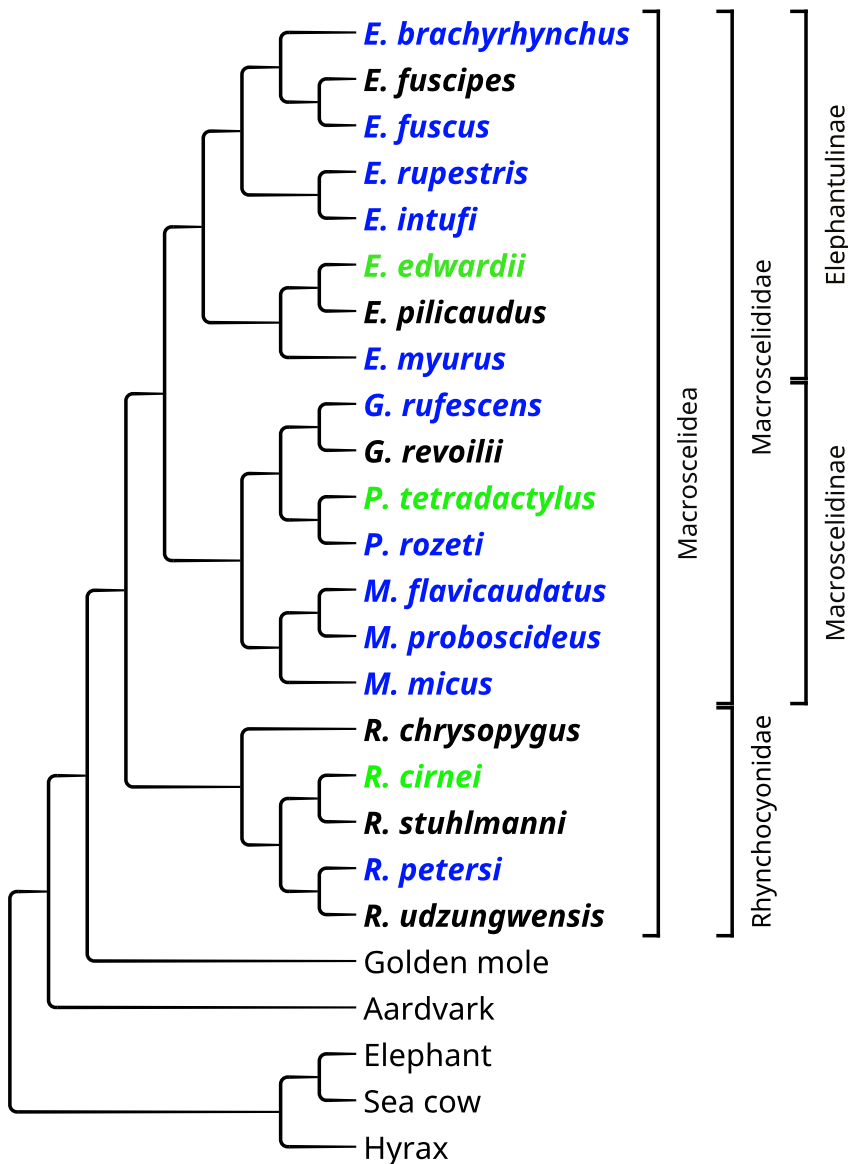
FIGURE 6.1: Simplified overview of assumed taxonomic relationships among all sengi species and other Afrotheria based on Hagemann et al. (2023). Green color indicates sengi species of which de-novo genomes were included in this study. Blue color indicates sengi species of which reference-based assembled genomes were included in this study. Black color indicates sengi species of which no genomic resources are available and which were thus not included in this study. Names on the right denote different phylogenetic entities of sengis.

Around 14Mb of one-to-one orthologous, protein-coding DNA allow us to confirm the assumed sengi phylogeny and detect potential gene-flow or incomplete lineage sorting (ILS). This comprehensive data set furthermore allowed us to narrow down the uncertainties around estimated node ages within sengis. We further tested

if window trees of the whole genomes show a higher amount of phylogenetic uncertainty or differences in topology compared to only the protein-coding fraction. In order to detect genes of evolutionary importance, we filter for genes that experienced either positive selection, gene-family expansion, gene-family contraction, or were lost along the evolution of sengis. We subsequently connect these genes to mammalian phenotypes and investigate if specific phenotypes can be found in all four filter categories across five phylogenetic clades, Macroscelidea as a whole, Rhynchocyonidae, Macroscelidea, and the two subfamilies of soft-furred sengis Elephantulinae and Macroscelidinae (Figure 6.1). Additionally, this study takes a closer look at the evolution of olfactory receptor proteins (ORs) within sengis and investigates the demographic history of the analyzed sengi species in comparison to other small African mammals. Besides substantially expanding the available genomic resources for sengis, this study reports precise, species-specific substitution rates and multiple candidate genes of potential importance for sengi evolution as well as potential phenotypic differences between different groups of sengis which might have been too cryptic to be observed previously.

6.3   Materials and methods

6.3.1   Samples

Genomes of fourteen different sengi species were analyzed during this study. One was retrieved from public databases (*Elephantulus edwardii*, GCA004027355.1), and thirteen were produced during this study. From the latter, *Petrodormus tetradactylus* (CAS29341) and *Rhynchocyon cirnei* (CAS29344) were assembled de-novo. The remaining eleven *Elephantulus brachyrhynchus* (CAS31689), *Elephantulus fuscus* (RS1240), *Elephantulus intufi* (ANG0160), *Elephantulus myurus* (RLMAR156), *Elephantulus rupestris* (CAS29696), *Galegeeska rufescens* (ER27111), *Macroscelides flavicaudatus* (CAS29703), *Macroscelides micus* (CAS29752), *Macroscelides proboscideus* (CAS28566), *Petrosaltator rozeti* (CAS27982), and *Rhynchocyon petersi* (RHY3717) assemblies were reference-based (Supplemental Table 1; Figure 6.1).

6.3.2   Extraction, library construction and sequencing

Genomic DNA from tissue samples for reference-based assemblies was extracted using a DNeasy Blood and Tissue kit (QIAGEN, Venlo Netherlands), sheared to a target size of 500bp using a Covaris S220 System (Covaris, Woburn, US-MA), and built into double-stranded, double-indexed Illumina sequencing libraries. High molecular weight DNA (HMW) for de-novo assemblies was extracted using a MagAttract HMW DNA kit (QIAGEN, Venlo Netherlands) and built into 10x Chromium Genomics libraries (10x Genomics, US-CA). All sequencing was conducted at a commercial sequencing facility (Novogene, London). Libraries were sequenced using 150bp paired-end reads on a NovaSeq6000 system (Illumina, San Diego, US-CA).

6.3.3   De-novo assemblies and repeat masking

Quality screening revealed that some reads had two 10x Chromium barcodes. Those reads were removed from the data set with BBMap (BBMap - Bushnell B. - sourceforge.net/projects/bbmap). Supernova v2.1 (Visendi, 2022), the assembler for 10x chromium linked reads, has a maximum number of input reads. Our raw reads surpassed this threshold. In order to improve assembly quality and contiguity, 10x barcode abundance within the data set was quantified with proc10xG (github.com/ucdavis-bioinformatics/proc10xG). Reads with barcodes which were present in less than 50 reads were subsequently removed to reduce read count and increase the ratio of reads per genomic fragment. Subsequently, Supernova v2.1

was run to assemble the two genomes with the maximal input of reads. Species-specific repeats were identified and masked with RepeatModeler v2.0.2a (repeat-masker.org/RepeatModeler/). All genomic assemblies and raw reads produced for this study were uploaded to public databases (in progress, Supplemental Table 1).

6.3.4 Reference based assemblies

Eleven reference-based assemblies were produced for this study. Adapters and reads shorter than 50bp were removed with Cutadapt v.3.4 (Martin, 2011). Bwa v.0.7.17-r1188 mem (Li and Durbin, 2009, 2010) was used to map the reads of each sample to its phylogenetically closest de-novo assembled reference (either *P. tetradactylus*, *E. edwardii*, or *R. cirnei*; Figure 6.1) with a minimum mapping quality of q = 30. Duplicates were marked with samtools v1.12 rmdup (Li et al., 2009) and subsequently removed. Consensus sequences were called with the samtools mpileup command.

6.3.5 Annotation

All assemblies were aligned to the human genome (GRCh38/hg38) with the Make-Lastz-Chains pipeline (github.com/hillerlab/make_lastz_chains, Kent et al., 2003; Kirilenko et al., 2023; Osipova et al., 2019; Suarez et al., 2017) and default settings. The resulting chained alignments were annotated with TOGA (Kirilenko et al., 2023) and the human genome (GRCh38/hg38) as reference. Isoform information of human genes was downloaded from Ensemble biomart (Cunningham et al., 2022) and added to the input data.

6.3.6 Additional resources

Genomes of *Chrysochloris asiatica* (GCF_000296735.1), *Loxodonta africana* (ftp://ftp.broadinstitute.org/pub/assemblies/mammals/elephant/loxAfr4/), *Orycteropus afer afer* (HLoryAfeAfe2, DNA Zoo Consortium), *Procavia capensis* (HLproCap4, DNA Zoo Consortium) and *Tricheus manatus* (HLtri ManLat2, DNA Zoo Consortium) were used as outgroups in this study. Human (GRCh38/hg38) reference-guided TOGA (Kirilenko et al., 2023) annotations of these genomes were downloaded (genome.senckenberg.de/download/TOGA/human_hg38_reference/) and integrated into the data set.

6.3.7   Phylogenomics

The respective longest transcript of each one-to-one ortholog of protein-coding genes of all 19 (with outgroups) or 14 (without outgroups) genomes were extracted. Genes that were not present in all samples were removed. Gene alignments were produced with MAFFT v7.480 (Katoh and Standley, 2013), concatenated with AMAS (Borowiec, 2016), and partitioned into 1st-2nd and 3rd codon positions for each gene. IQ-Tree v2.0.3 (Minh et al., 2020) was used to produce gene trees and merge partitions with PartitionFinder v2.1.1 (Lanfear et al., 2016). To estimate the species tree, gene trees from IQ-Tree were analyzed under default settings with ASTRAL v5.7.8 (Zhang et al., 2018). Gene tree discordance was estimated and visualized with DiscoVista (Sayyari et al., 2018).

To test for potentially different phylogenetic signals of coding and non-coding parts of the genomes, WindowTrees (https://github.com/achimklittich/WindowTrees, Hempel et al., 2021) were run. Alignment chains with pairwise alignments to the human genome from the annotation step of all fourteen sengi genomes were converted to sorted .maf format with the UCSC-tools package (http://genome.ucsc.edu) for each human autosome. Overlapping regions of .maf files were removed with the tool single_cov2 (UCSC-tools). All pairwise alignments were then converted to a multi-sequence alignment with the program roast (UCSC-tools), using the human genome as a reference and the species tree from the ASTRAL analysis step as a guiding tree. Multi-sequence alignments (MSAs) in .maf format were subsequently converted to fasta format with maf2fasta (UCSC-tools). Window trees of these MSAs were calculated for 10,000bp windows, with a distance of 10,000bp between windows and a maximum of 50% Ns per window allowed. Temporal calibration of the species tree was conducted with MCMCtree and BASEML from the PAML software package for phylogenetic analysis (Yang, 2007). The previously created MSA of one-to-one orthologous protein-coding genes was used, as well as the partitioning of IQ-Tree/PartitionFinder. To speed up the process, the approximate likelihood calculation approach (Reis and Yang, 2011) was conducted. The species tree gained through the ASTRAL analysis was used as the input tree. As MCMCtree does not offer the usage of a local random clock, which was previously shown to be the most adequate clock model for sengis if analyzed in conjunction with outgroups and outgroup fossil priors (Hagemann et al., 2023), a global clock and HKY85 as the substitution model for each partition was used. *Oligorhynchocyon songwensis* from the Nsungwe Formation in the Rukwa Rift Basin of Tanzania (25.2 million years old; Stevens et al., 2022) was used to calibrate the macroscelidean crown node. Based on previous findings, no outgroup sequences or additional fossil calibration points from outside Macroscelidea were used (Hagemann et al., 2023. MCMCtree was run

10 times for 1,000,000 iterations after a burn-in of 20,000 iterations. Convergence of each run was confirmed with Tracer v1.7.2 (ESS larger than 200, andrew et al., 2018). MCMC chains of all ten runs were merged before producing the final temporally calibrated tree with BASEML.

6.3.8 Demography

Sex chromosomes/scaffolds within sengis were identified with SATC (Nursyifa et al., 2022), by normalized mapping depth, and subsequently removed from the assemblies. Trimmed raw reads were mapped to their respective assemblies with Bwa v.0.7.17-r1188 mem (Li and Durbin, 2009, 2010) and minimum mapping quality of 30. Duplicates were removed with samtools rmdup. Diploid genomes with a minimum read depth equal to a third of average read depth, a maximum read depth equal to twice the average read depth, and a minimum mapping quality of 30 were called. Pairwise sequential Markovian coalescent (PSMCs; Li and Durbin, 2011) were run with 100 bootstrap replicates and the following parameters: -N25 -t15 -r5 -b -p "4+25x2+4+6". Results were plotted with psmc_plot.pl. Generation times were taken from (Soria et al., 2021).

Species-specific mutation rates were calculated from the above-produced window trees by summing the branch lengths from the sengi MRCA to each terminal and dividing it by the mean age of the sengi crown node, determined by the MCMCtree step (30.8624 mya) for each window tree and averaging over the results.

To investigate if sengi's demographic past differed from other small African mammals, we also produced PSMCs for the Cairo spiny mouse (*Acomys russatus*, GCA_903995435.1), Percival's spiny mouse (*Acomys percivali*, GCA_907169655.1), Kemp's spiny mouse (*Acomys kempi*, GCA_907164505.1), the Nile rat (*Arvicanthis niloticus*, GCA_011762505.1), the lesser jerboa (*Jaculus jaculus*, GCA_020740685.1), the cape ground squirrel (*Xerus inauris*, GCA_004024805.1), and the unstriped ground squirrel (*Xerus rutilus*, GCA_028644305.1). All autosomal chromosome assemblies were downloaded from NCBI for each of these species. Raw reads from the Sequence Read Archive were mapped to the respective assemblies, and PSMCs were produced as described above. Inverse instantaneous coalescence rates were interpreted as a proxy for effective population sizes, although they can be affected by selection or nonrandom mating (Mather et al., 2020).

6.3.9   Adaptive genomic data set

Genes either identified as "clearly lost" (L), as one-to-many (gene family expansion) or many-to-one (gene family contraction) with regard to the human genome annotation (see (Kirilenko et al., 2023) and TOGA documentation for more information) were extracted from the TOGA annotation of all genomes (sengis and outgroups). To only identify sengi-specific patterns of gene evolution (not Afrotherian-specific ones), all genes which were classified as one-to-many, many-to-one, or lost in any of the outgroup genomes were removed from the data set. This filtering step should also minimize the possibility of false-positive one-to-many categorization of genes which experienced gene family contraction in only humans and therefore appear to be expanded in Afrotherians.

To investigate genes in specific phylogenetic clades within sengis and account for the potential incompleteness of the reference-based assemblies, the following filtering was applied: for a gene to be classified as "lost" (or expanded or contracted, respectively) it had to be classified as "lost" in all de-novo assembled genomes of the respective clade and was not allowed to be classified as "lost" in any sample outside of the clade of interest (de-novo assembled genomes and shotgun genomes). However, it was not necessary that the gene was classified as "lost" in the shotgun genomes within the respective clade of interest. Although this filtering appears as the most reasonable one for the here presented data, it has to be mentioned that it bears the risk that a gene is classified as e.g. "lost" in a lower phylogenetic clade because an outgroup de-novo genome assembly is missing the respective regions due to assembly quality.

Applying the above restrictions, we filtered for genes for the three categories lost, expanded (one-to-many), and contracted (many-to-one) in five phylogenetic clades: all sengis (Macroscelidea), giant sengis (Rhynchocyonidae), soft-furred sengis (Macroscelidae), Elephantulinae, and Macroscelidinae. In addition, we tested which genes from the one-to-one orthologs data set (with outgroups) were positively selected on the branch leading to the five clades. paPAML (Steffen et al., 2022), a parallelized version of PAML (Yang, 2007), was used to detect genes under positive selection using the branch-site model, with $p < 0.05$ and default settings. In order to minimize the possibility of type I errors (false positives), we subsequently corrected for multiple testing, using the Benjamini-Hochberg procedure in the python module statsmodels (Seabold and Perktold, 2010), and an alpha of 0.05. Taken together, for each of the five sengi clades in focus, four sets of genes were extracted which were either exclusively positively selected, lost, expanded, or contracted, totaling to 20 gene sets, which were used for further processing.

6.3.10 Gene set enrichment tests

For each of the 20 gene sets from the previous step, a Gene Ontology (GO)-term enrichment analysis was conducted via the Gene Ontology network (Aleksander et al., 2023; Ashburner et al., 2000). Homo sapiens was used as the reference list, and enrichment was tested for the category "GO biological process". Results were considered significant if q < 0.05 after correcting for multiple testing with the Benjamini-Hochberg procedure.

Additionally, the enrichment of mammalian phenotypes in the 20 gene sets was tested with the python package GSEApy (Fang et al., 2023). The mammalian phenotype database from (Marcovitz et al., 2019) based on the Mouse Genome Informatics database (MGI) (Blake et al., 2021) was used.

After correcting for multiple testing as described above, phenotypes significantly enriched for each category within a clade were compared to identify similar/same phenotypes enriched within each category. However, after correction for multiple testing, nearly no significantly enriched phenotype was found. We therefore decided to use a different approach with less strict error correction for multiple testing (see discussion). Enrichment analysis for mammalian phenotypes and the 20 sets of genes were repeated as described above. No cut-off value for q after correction for multiple testing was defined, but enriched phenotypes were sorted by q values, and up to 1500 phenotypes with the lowest q values were kept for further processing. To qualitatively account for randomly enriched mammalian phenotypes, only phenotypes which were detected in each of the four categories for a single phylogenetic clade were considered to play an important role in sengi evolution.

Enrichment analysis of genes that have undergone multiplication (one-to-many) is bent towards insignificant results. For example, if the gene PRR13 is multiplied 10 times within all giant-sengis, this might point towards a special adaptive function of the respective gene. However, when conducting an enrichment analysis, the gene would only be counted once, not 10 times. As we are not aware of a meaningful way to correct for this issue, genes that experienced extensive multiplication (> 3 copies) in each member of the five analyzed phylogenetic clades were investigated individually by associating potentially affected phenotypes via the GeneCards database (Stelzer et al., 2016) and the Human Disease Ontology database (Schriml et al., 2019).

6.3.11 Curation and Comparative Genomics Analysis of Olfactory Receptor Repertoires.

We employed a multi-step approach to identify Olfactory receptors (ORs) in the 14 Macroscelidea taxa. Initially, we obtained all reviewed ORs from the SwissProt

database and all previously published mammalian OR sequences from earlier studies (PMIDs: 17684554, 14507991, 17166524). All recovered sequences were clustered using the BLASTCLUST program, ensuring no cluster had more than a 40% identity threshold. Each cluster was aligned using the MAFFT program (PMID: 23329690), with alignments carefully inspected for completeness and accuracy, and separate HMM profiles were constructed using the HMMER software. Likewise, HMM profiles specific to the ORs (7TM_4) were downloaded from the Pfam database. This resulted in a prudent list of distinct HMM profiles utilized as search seeds for HMM-based searches across all extracted proteomes from Macroscelidea. Secondly, reviewed ORs were used as search seeds for TBLASTN searches against all 14 Macroscelidea genomes. All recovered receptors from HMM searches against the proteome and all translated open reading frames obtained from the TBLASTN searches were subjected to a four-step validation process: (i) reciprocal BLAST searches were performed against the Refseq and Swissprot databases, and only the sequences that recovered ORs as their top hits were kept; (ii) putative ORs were inspected for completeness by predicting the number of transmembrane segments using Phobius and DeepTMHMM programs (PMID: 17483518, https://github.com/cbligaard/DeepTM) (iii) all predicted ORs were subject to phylogenetic analysis with the non-ORs (other Class-A GPCRs) of the same genome and verified whether the predicted ORs formed a separate and distinct cluster; (iv) all identified OR encoding genes were cross-verified with TOGA annotation pipeline to ensure accuracy of the identified gene-loci and eliminate any mapping bias. Lastly, to eliminate any redundancy and inflation in the number of predicted ORs, all sequences that passed the above criteria within each taxon were clustered at a 98% identity threshold using CD-HIT (PMID: 23060610). This resulted in the final list of ORs in all 14 Macroscelidea, as shown in Figure 6.4, and was utilized for further comparative genomics analysis. To gain insights into species-specific adaptations and contractions, we conducted a comparative genomics analysis of the OR repertoires across all the analyzed Macroscelidea species, using the OrthoFinder program with the following parameters. We utilized BLAST as the sequence search tool, MAFFT for multiple sequence alignment, IQTREE for tree inference, STAG for species tree inference, and a default inflation parameter of 1.5 for the MCL (Markov clustering algorithm) clustering. The results from the OrthoFinder program included comprehensive reconstruction of orthogroups, orthologs, gene trees, resolved gene trees, rooted species trees, gene duplication events, and comprehensive comparative genomic statistics. These included the number of one-to-one orthologs between each pair of species, the count of orthologs in one-to-many relationships (indicating gene duplication in one of the two lineages post-speciation), and the number of orthologs in many-to-many relationships for each species pair

(reflecting gene duplication events in both lineages post-speciation). The inferred data sets and matrices were visualized and plotted using the ggplot2 package in R language.

6.4   Results

6.4.1   New genomic resources

De-novo assemblies from 10x chromium linked reads of CAS29341 (*P. tetradactylus*) and CAS29344 (*R. cirnei*) with Supernova (Visendi, 2022) resulted in fairly complete but fragmented genomic assemblies. The N50 metric was 503 Kbp and 770Kbp for CAS29341 and CAS29344, respectively. The L50 metric was 2167 and 1926, respectively. In CAS29341, 83% of mammalian BUSCO genes were found, and 90.3% in CAS29344. Substantially differing genome sizes were detected with 3.96Gbp (*P. tetradactylus*) and 5.47Gbp (*R. cirnei*) between soft-furred and giant-sengis. The RefSeq genome of *E. edwardii* has a size of 4.07Gbp. The mapping coverage of the reference-based assemblies was between 13x-41x, and mammalian BUSCOs for these assemblies were between 78-94% complete (Supplemental Table 1). Annotation with TOGA resulted in between 18,497 and 18,939 protein-coding/pseudogenized orthology components for each genome (one-to-one, one-to-zero, many-to-one, one-to-many, many-to-many; Figure 6.2). The number of one-to-one orthologs with the human genome ranged between 12,741 (*G. rufescens*) and 15,204 (*E. myurus*) . The data set without outgroups consisted of 8,973 one-to-one protein-coding orthologs present in all samples with a total length of 14,995,958 bp. The data set including outgroups consisted of 8,024 such orthologs present in all 19 species and had a total length of 13,754,952 bp. This study, therefore, fundamentally extended the available genomic resources of Macroscelideans for future research.
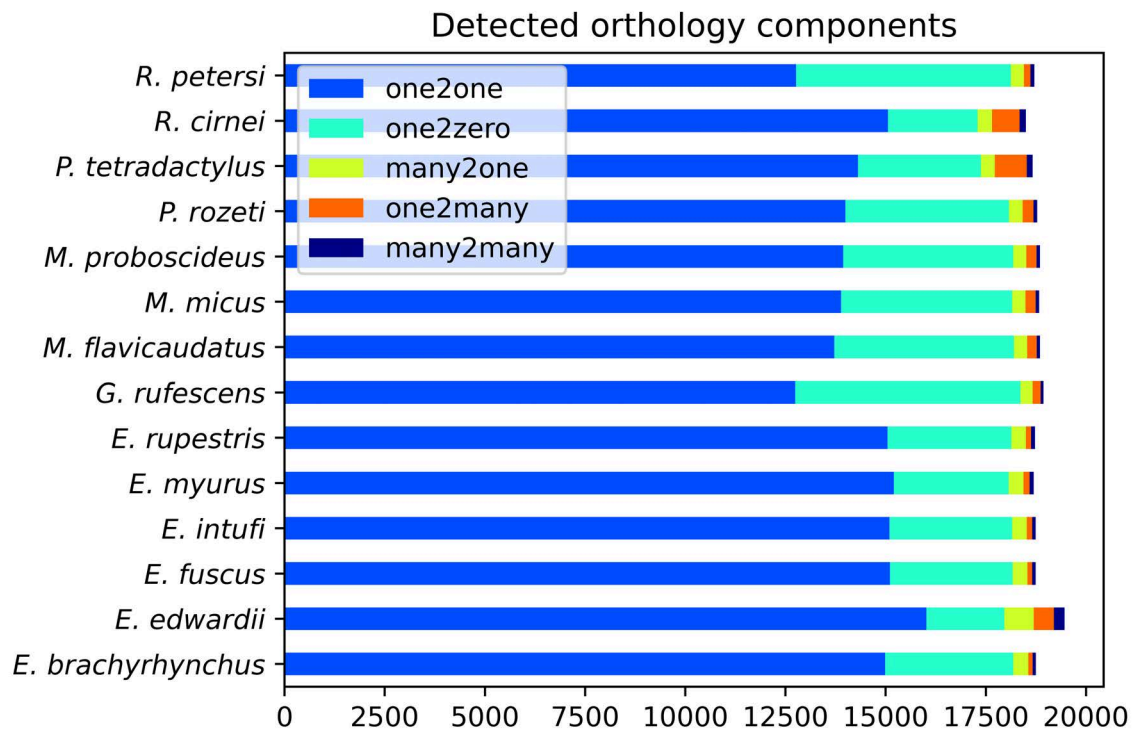
FIGURE 6.2: Detected protein-coding/pseudogenized genes by TOGA
and their ortholog relationships in regard to the human genome

6.4.2 Phylogenomics and gene tree discordance

The 8,973 (8,024 with outgroups) gene trees produced with IQ-Tree showed a very
low level of gene tree discordance. Local posterior probabilities of all quadripar-
titions consistently supported a single topology above the other two alternatives,
indicating the absence of extensive ILS in sengis as well as post-speciation gene-flow.
Accordingly, the species tree was well resolved (Figure 6.3; Supplemental Figure 1).
The highest amount of gene tree discordance was detected on the branch leading
to Macroscelidinae, whereby the species tree topology was still supported by 71%
of all gene trees. Mostly, branches within Macroscelidinae showed more gene tree
discordance than branches outside of this clade (Figure 6.3; Supplemental Figure 1).
This result was further supported by the window tree analysis, investigating topo-
logical differences of window trees across the entire genome without being restricted
to protein-coding regions. Out of 7562 window trees, 98% (7398) supported a single
topology, the same as that identified as species tree topology by ASTRAL. In total,
2% (164) window trees with 58 alternative topologies were found. However, 47 of
those topologies were solely supported by 1-2 window trees and therefore considered
as noise (See Supplemental Table 2 for all window trees). Two alternate topologies
received a somewhat higher support, both regarding the relationship between the

genera *Petrosaltator*, *Petrodromus*, and *Galegeeska*. In contrast to the species tree from ASTRAL and the vast majority of the window trees, which support a sister relationship between *Petrosaltator* and *Petrodromus*, a sister relationship between *Galegeeska* and *Petrosaltator* was supported by 30 window trees and a sister relationship between *Galegeeska* and *Petrodromus* was supported by 26 of such trees (Supplemental Table 2). PartitionFinder recommended 432 partitions for the 8,973 genes. The age of the sengi crown-node was estimated to be 30.9 mya (26.4-34.8, 95%HPD). The age of soft-furred sengis was estimated to be 9.5 mya (8.1-10.7, 95%HPD). Macroscelidinae were estimated to be 8.2 my (7-9.2 95%HPD) old, similar to Elephantulinae with 7.5 my (6.4-8.5 95%HPD). The age of the split between the two giant sengis included was estimated to be 374,000 years (318,000-421,000 95%HPD; Figure 6.3). Substitution rates estimated from window trees with the species tree topology were between 4.77×10^{-9} [substitutions per site per million years] (*E. edwardii*) and 5.52×10^{-9} [substitutions per site per million years] (*R. cirnei*) (Supplemental Table 3). Substitution rates of only the protein-coding part of the genome were roughly half that fast with $2.13\text{-}3.02 \times 10^{-9}$ [substitutions per site per million years].
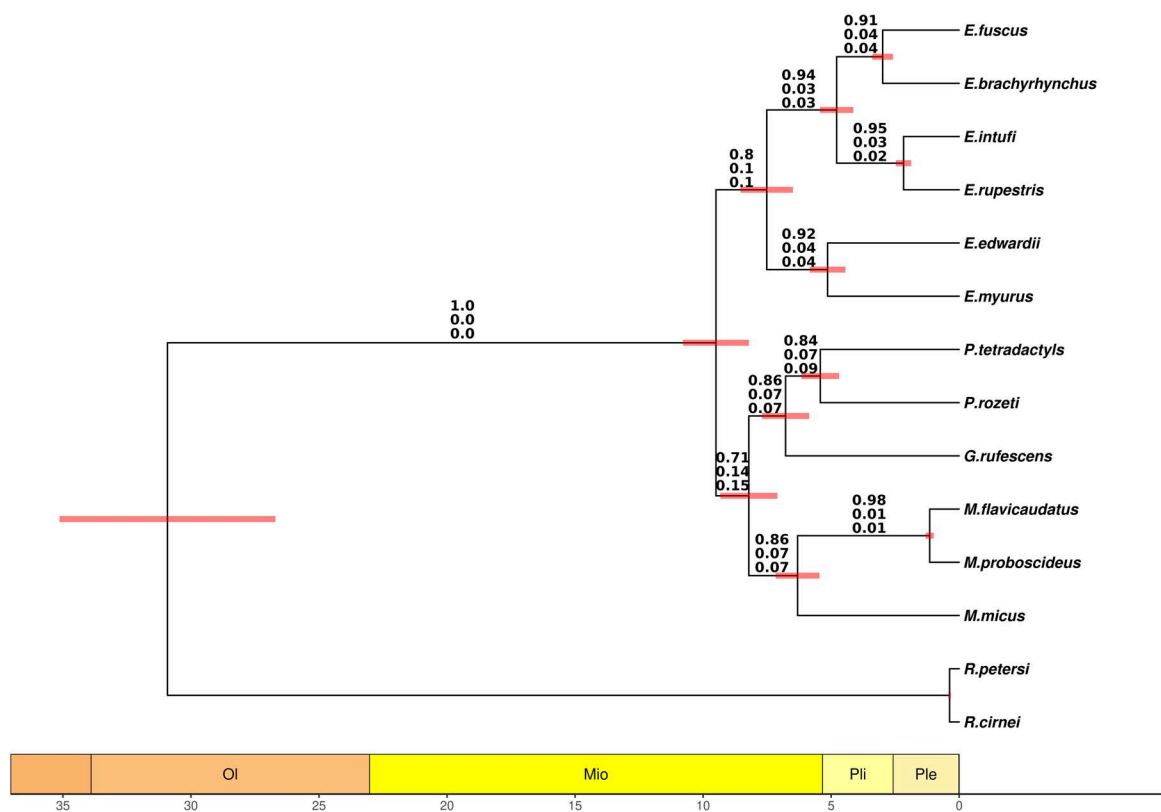
FIGURE 6.3: Temporal calibrated species tree of Macroscelidea based on 8,973 one-to-one orthologous protein-coding genes. Numbers above branches show local posterior probabilities of quadripartitions inferred by ASTRAL and DiscoVista. Top numbers per branch indicate support for the here displayed phylogenetic relationships per quadripartition, while both lower numbers indicate the support for the other two alternative topologies.

6.4.3 Demographic past

PSMCs revealed no overall sengi-specific demographic history compared to other small African mammals. On the contrary, *E. edwardii*, *R. petersi*, and *P. rozeti* showed a similar demographic history as *X. rutilus*, *X. inauris*, and *A. niloticus*. *Petrodromus tetradactylus* and *E. intufi* had similar PSMC plots as *J. jaculus* and *A. russatus*. However, eight out of the twelve soft-furred sengis analyzed during this study showed a decline in effective population size between 1,000,000-700,000 years ago (Supplemental Figure 2). Lowest recent population sizes were found in the two *Rhynchocyon species*, *P. rozeti*, and *E. edwardii*, with all showing effective population sizes below $5x10^4$. This might be to some extent explained by the comparatively small estimated species ranges of *edwardii*, *petersi*, and *cirnei*. However, the range of *P. rozeti* is neither particularly small nor large for sengis. All of these four species experienced a more or less steady decline in population size over the past approx. 1.000.000 years

(Supplemental Figure 2). The largest recent effective population sizes of sengis were detected in *P. tetradactylus* and *E. intufi* with around 4×10^5. *P. tetradactylus* has one of the largest estimated ranges of all sengis (IUCN Red List; Rathbun, 2015). *E. intufi*'s range is similar to *P. rozeti*, neither particularly small nor large for sengis. Both species experienced a rise in effective population size between 800,000-500,000 years ago.

6.4.4   Olfactory receptor protein evolution

When analyzing only sengi genomes, the number of shared one-to-one orthologs among species follows only to some extent their phylogenetic relationships (Figure 6.4). In the subfamily Elephantulinae, the number of one-to-one orthologs in pairwise comparisons ranges from 357-434 ORs and 342 in the genus *Rhynchocyon*. The subfamily Macroscelidinae did not show such a large number of shared ORs, although the phylogenetic distances between the species of this clade are similar to those in Elephantulinae (Figure 6.3 and Figure 6.4). Within Macroscelidinae, only 224-380 ORs were detected in pairwise comparisons. Within this family, the three *Macroscelides* species included share an amount of ORs in accordance with their phylogenetic distance. However, no such pattern was found for the three species *G. rufescens*, *P. tetradactylus*, *P. rozeti*. Especially *P. tetradactylus*, although being the reference for all other genomes within Macroscelidinae, shares a comparatively small number of ORs with the other species from this clade (224-254; Figure 6.4).

OR genes in one-to-one relationships between species
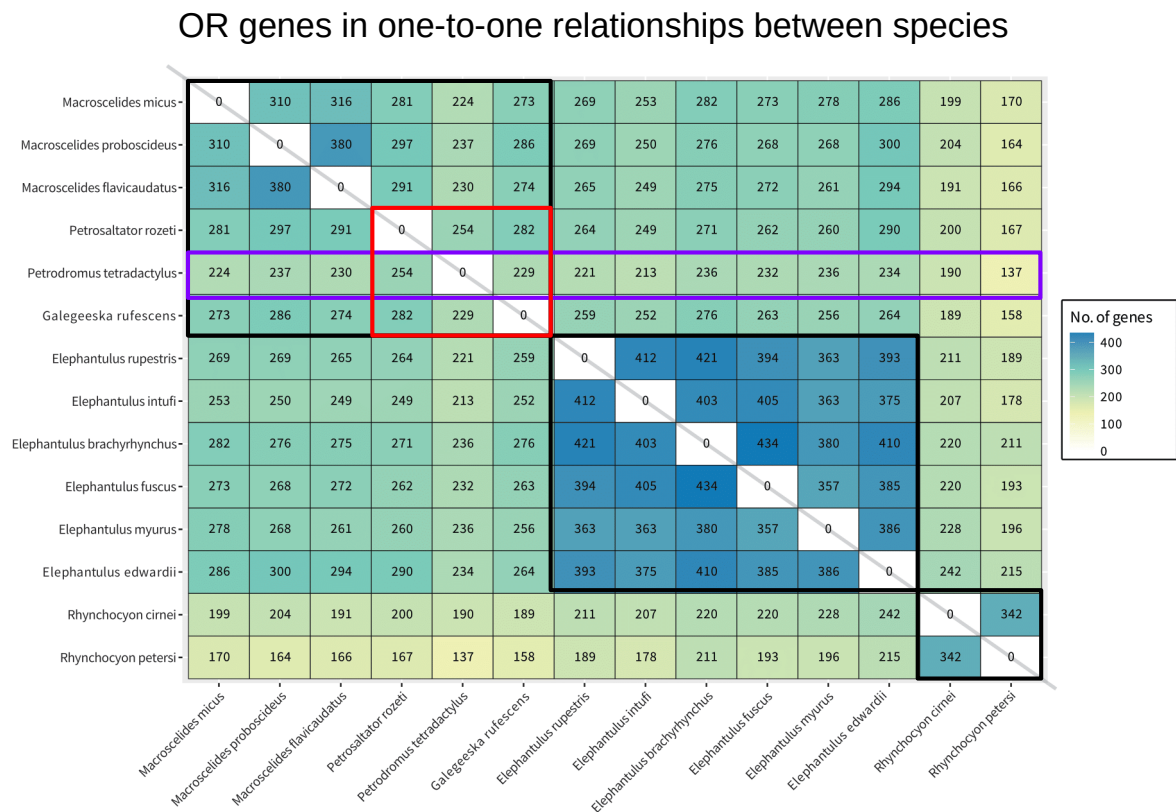
FIGURE 6.4: Olfactory receptor (OR) protein genes in one-to-one orthologous relationships across all sengi species analyzed. Black boxes indicate phylogenetic entities. Uppermost: Macroscelidinae; middle: Elephantulinae; bottom: Rhynchocyonidae. Red box highlights the low amount of shared ORs between the three species *P. rozeti*, *P. tetradactylus* and *G. rufescens*. Purple box highlights the low amount of shared ORs of *P. tetradactylus* with any sengi species, besides being the assembly reference of all other Macroscelidinae genomes.

6.4.5 Genes under special evolutionary modes and associated phenotypes

In order to detect physiological traits of importance for sengi evolution, lists of genes for the five phylogenetic clades (Macroscelidea, Macroscelididae, Rhynchocyonidae, Elephantulinae, Macroscelidinae) which were either positively selected, lost or experienced gene-family contraction or expansions were tested for enriched GO-terms or associated mammalian phenotypes. Enrichment analysis for both GO-terms and mammalian phenotypes in the 20 sets of genes (four per phylogenetic clade) resulted in only very few significantly enriched terms, after correction for multiple testing (Supplemental Table 4 and 5). We therefore linked those genes to their associated mammalian phenotypes (without enrichment test) and subsequently tested if shared phenotypes among the four modes of evolution exist, within a single phylogenetic

clade. No phenotypic terms shared among sengis as a whole were found, but for Macroscelididae (13), Rhynchocyonidae (1), Elephantulinae (44), and Macroscelidinae (8; Figure 6.5; Supplemental Table 6). Within Macroscelididae, these terms were associated with the nervous system, the hematopoietic system, the immune system, and locomotion. The single term found within Rhynchocyonidae affected the hematopoietic system. In Macroscelidinae, half the terms were associated with water metabolism (balance?), the remaining ones with locomotion, homeostasis, and again the hematopoietic system. Most terms were found within Elephantulinae, affecting the nervous system, once more the hematopoietic system, hormones and homeostasis, locomotion, and vision. Ten terms affecting morphology like e.g. "decreased body size", "abnormal long bone morphology", and "abnormal myeloid cell morphology" were found. Another ten terms were associated with reproduction, including "prenatal lethality" or "embryonic lethality between implantation and placentation" (see Supplemental Table 6 for full lists).
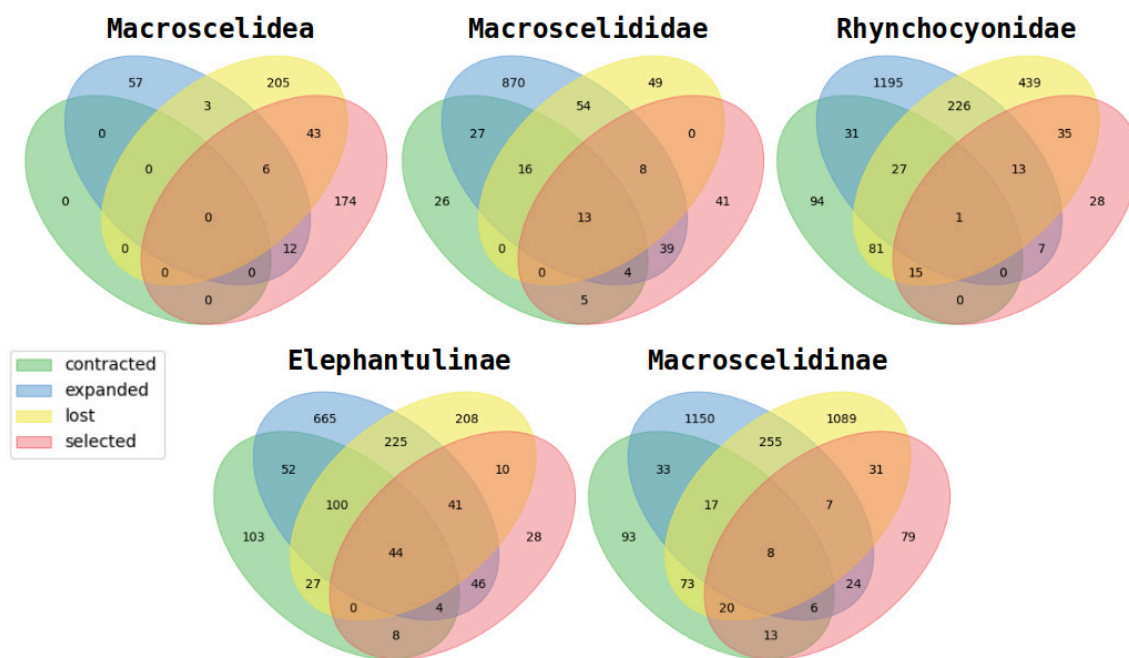
FIGURE 6.5: Amounts of shared phenotypic traits associated to genes which were either positively selected, experiences gene-family expansion or contraction or were lost for a specific phylogenetic clade within sengis.

6.4.6 Extensively multiplied genes

In addition, we extracted genes which experienced extensive multiplications (> 3 duplication) in all representatives of their phylogenetic group (See Supplemental Table 7 for full lists and copy numbers). We found that a single gene (ENSG00000153446)

was extensively multiplied in every sengi genome analyzed here, with ten to 32 (average fourteen) copies present in each genome, suggesting an important role of ENSG00000153446 in sengi evolution. In giant sengis, we found a total of 35 genes with extensive multiplications. Among the ones most heavily duplicated were ENSG00000065809 (eleven copies per genome) and ENSG00000151135 (nine copies per genome). Within soft-furred sengis, three genes with multiple copies in each representative were found, ENSG00000135346 (six to eight copies), ENSG00000170835 (four to six copies), and ENSG00000187555 (four to six copies). Within Macroscelidinae, eighteen genes with multiple copy numbers were found. The most extensively multiplied genes were ENSG00000065534 (twelve copies), ENSG00000141837 (ten copies), and ENSG00000186868 (ten copies). Within Elephantulinae, we only detected three such genes, ENSG00000132122 (four to six copies), ENSG00000137501 (six to seven copies), and ENSG00000204371 (six to eight copies).

6.5   Discussion

The mammalian order Macroscelidea is a remarkable group of Afrotherian mammals with a poorly known evolutionary history, restricted to the African continent. By generating new genomic resources for thirteen out of the twenty currently recognized sengis species, this study laid the foundation for a better understanding of sengis' evolutionary past through genomics. The application of a vast variety of analysis revealed multiple insights, from basic taxonomy to functional implications of potential adaptive genes.

In the age of genomics, it is not uncommon that mammalian evolutionary studies reveal a considerable amount of gene tree discordance, oftentimes explained by post-speciation gene-flow, incomplete lineage sorting or both (e.g., Árnason et al., 2018; Barlow et al., 2018; Fontsere et al., 2019). Sengis display a high level of morphological resemblance among the species; some occur in sympatric distributions inhabiting very similar niches. Besides the Saharan desert separating *P. rozeti* from other sengis, no "hard" geographic barrier exists between the species' distribution (IUCN Red List; Perrin and Rathbun, 2013; Rathbun and Dumbacher, 2015). Under these circumstances and the usual amount of gene-flow observed in mammals, the detection of gene tree discordance in sengis would be no surprise. However, post-speciation gene-flow and ILS seemingly played only a negligible role during their evolution. Classical bifurcating speciation processes seem to have happened in the evolutionary past of sengis. This finding somewhat adds another enigmatic aspect to our knowledge of this remarkable clade of mammals. Future research should further investigate this phenomenon and test to which extent this picture can also be found among populations of a single sengi species.

Although nearly no gene tree discordance was detected overall, within Macroscelidinae and especially regarding the relationships between *P. tetradactylus*, *P. rozeti*, and *G. rufescens*, a comparatively higher amount was found when protein-coding parts of the genome as well as whole genomes were analyzed. Remarkably, this result, to some extent, echoes the past uncertainty of placing the latter two species within the sengi phylogeny, based on morphological evidence. Both species, *P. rozeti* and *G. rufescens*, were long assumed to belong to the genus *Elephantulus* and were only reassigned to their present genera based on molecular evidence (Douady et al., 2003; Krásová et al., 2021). It also reflects the uncertainty of the exact branching order of the three species recognized when using only a small number of nuclear genes (Hagemann et al., 2023; Krásová et al., 2021). While the current distributions of *G. rufescens* and *P. tetradactylus* are in close spatial proximity and even overlap in Tanzania, both are separated from *P. rozeti* by the Saharan desert (IUCN Red List;

Rathbun and Dumbacher, 2015). However, in contrast to the hyper-arid condition of the Sahara today, during multiple wet phases over the past 194ky, the Sahara was vegetated and harbored forests, grasslands, and permanent lakes (Castañeda et al., 2009). The biogeographic barrier Sahara, at least for some species, was thus more permeable in the past, potentially affecting the distribution of the three species and allowing gene-flow among them. However, further investigations are needed to closer examine the gene-flow patterns among the three species. These studies should also include *G. revoilii*, the only remaining member of this clade for which no genomic information is available at the moment.

The evolutionary age of sengis was a long-standing question which only recently was resolved (Hagemann et al., 2023; Heritage et al., 2020; Krásová et al., 2021). However, previous temporal calibration studies included only a few nuclear loci. This limited amount of data resulted in large node age uncertainties, plus being not informative enough to fully resolve all nodes. The here generated data set of 8,973 protein-coding orthologs enabled us to surpass these shortcomings and produce a high confidence temporal calibrated phylogeny of sengis with substantially smaller node age uncertainties. However, extensive amounts of data also bear their challenges. As previously shown, it is desirable to use multiple fossil priors to calibrate multiple nodes when generating a temporally calibrated tree (Bibi, 2013). However, for sengis, this implies including outgroups, as the sengi fossil record allows only reliably calibrating the sengi crown-node (Holroyd, 2010; Stevens et al., 2022). Furthermore, it is mandatory for obtaining reliable results that the right clock model is chosen (Bibi, 2013; Hagemann et al., 2023). For Macroscelidea, a random local clock is recommended when including other afrotheres as outgroups (Hagemann et al., 2023). Unfortunately, this clock model is not supported by MCMCtree, a software capable of processing large amounts of data in a reasonable time. However, previous research also showed that the clock model selection is less critical to producing reliable calibrations if no outgroups (and outgroup fossil priors) are included (ibid.). We, therefore, did not include outgroups and multiple fossil priors into the temporal calibration when using our genome-scale data set for obtaining the dated phylogeny. The here-produced temporal calibration dates the split between the two families Macroscelididae and Rhynchocyonidae to the early Oligocene. One of the major fossil sites for sengis in southern Africa is Sperrgebiet, Namibia. However, the ages of the Paleogene localities are disputed among scientists. While multiple publications assume older ages (e.g., Pickford and Senut, 2008; Senut and Pickford, 2021), others advocate younger ones (e.g., Bronner et al., 2023; Marivaux et al., 2014; Sallam and Seiffert, 2016). If assuming the younger ages, the time of the Macroscelidean

crow-split into the two families was likely part of a major sengi radiation that resulted in high lineage diversity in eastern (Butler, 1995; Stevens et al., 2022) and southern Africa (Pickford and Senut, 2008; Senut and Pickford, 2021) during the middle and late Oligocene as well as the earliest Miocene. Besides the members of the two extant families, all this diversity was subsequently lost and is only known from the fossil record (Butler, 1995; Pickford and Senut, 2008; Senut and Pickford, 2021). If this loss in diversity is a result of competition with other (Laurasian) immigrants, climatic changes, or both remains unclear. The little morphological changes observed in the fossil record might argue against enhanced competition as no specialization/morphological evolution can be observed (Holroyd, 2010). The radiation of soft-furred sengis is estimated to have started towards the late Miocene. These dates are in accordance with previously published ones based on nuclear DNA and only a few marker genes (Hagemann et al., 2023; Krásová et al., 2021). They, therefore, are consistent with the proposed climate cycle and resulting forest fragmentation-driven evolution of sengis, as proposed by Krásová et al., 2021. A climate-driven evolution might be further supported by the absence of gene-flow, which appears somewhat more plausible under conditions of arising gene-flow barriers through climate-driven habitat fragmentation, compared to a competition-driven evolution through immigrating Laurasian mammals. However, to which amounts sengi evolution was shaped by climatic events and competition through other mammals need to be elaborated in future research.

Further insights about the simultaneous evolution of sengis and other small African mammals possibly could be gained through a comparison of their demographic past. However, the modeling of the demographic histories of sengis in conjunction with other African mammals, presumably occupying similar niches, did not reveal a sengi-specific pattern, neither by phylogenetic proximity nor by other life history traits like habitat type usage. Depending on the mutation rate, population size, and generation time, PSMC can only model demographic changes of the more recent past before all alleles coalesced, for sengis 1my (Li and Durbin, 2011; Mather et al., 2020). As shown in the phylogenomic part of this study, the age of sister species splits mostly surpass this time frame considerably. Thus, PSMCs are no optimal tool to detect shared demographic events among sengis, as the PSMC plots of e.g. sister species will not have a common origin, which would increase the resilience of the method. However, the split between the two giant sengis is dated to 318,000-421,000 years ago, a period where both PSMC graphs overlap, which provides some confidence in the PSMC graphs produced here (Supplemental Figure 2 and 3). Apart from *E. rupestris*, *P. tetradactylus*, and *E. intufi*, all sengis have rather low effective population sizes (Ne) and experienced a decline in Ne over the past approx. 500ky. If coupling

Ne to the vulnerability to extinction, our results indicate that the extinction risk of *R. petersi*, *R. cirnei*, *E. edwardii*, and *P. rozeti* should potentially be (re)evaluated, but see (Kyriazis et al., 2021). Most sengis are not assumed to be particularly threatened and all species analyzed here are considered either "Least Concern" or "Data Deficient" on the IUCN red list. However, it has been noted that the status of all giant-sengis should be of concern as populations are decreasing and their habitats are more affected by logging, agriculture, and urban development (Rathbun, 2009b, sengis.org). Soft-furred sengis, in comparison, mostly dwell in habitats less affected by these human activities (Heritage, 2018; Rathbun, 2009b, sengis.org).

Besides the phylogenomic analysis so far described, this study also investigated more functional aspects about sengi evolution. The trunk-like, mobile snouts of sengis indicate an outstanding importance of the olfactory sense for this mammalian clade. However, the analysis of OR evolution and comparisons across clades is complicated by different phylogenetic distances, the number of representatives per clade as well as differences in quality between de-novo and references based assemblies. Therefore, no cross Afrotherian comparison was conducted, and the authors recommend to not focus on specific numbers of the OR analysis carried out here but rather at the trends of the analyses. In accordance with the phylogenetic insights, the OR analysis pointed towards a different mode of evolution of the three species *G. rufescens*, *P. tetradactylus*, and *P. rozeti*. While the phylogenetic clades of Rhynchocyonidae and Elephantulinae show a high number of shared ORs among their members, within Macroscelidinae such a picture could only to some extent be observed among species of the genus *Macroscelides*, but not among the previously mentioned three (Figure 6.4). This result was most clear for *P. tetradactylus*, the only species within Macroscelidinae occupying forested or wooded habitats with a closed canopy and higher precipitation compared to all other members of this group (Corbet and Hanks, 1968; Jennings and Rathbun, 2001; Perrin and Rathbun, 2013).

In addition to the OR analysis, this study took a number of different approaches in order to identify genes of potential adaptive value and the phenotypes shaped by them. With the application of NGS methods, especially in diagnostics, it has become apparent that most phenotypes are not shaped by a single gene but frequently by > 100 genes. Furthermore, a single gene does not only shape a single phenotype, but multiple (e.g., Gannamani et al., 2021). Taking this into consideration, it appears plausible that the underlying genes of a single phenotype are affected in different ways and with different consequences by evolutionary forces, optimizing a specific phenotype to improve its adaptive value for a specific environment. We therefore identified genes which experienced positive selection, gene-family expansion, gene-family contraction, or those which have been lost during sengi evolution for five

phylogenetic entities within sengis. These genes were subsequently linked to mammalian phenotypes (based on the mouse genome informatics project; MGI) and tested if any phenotypes could be detected in each of the four categories within a single phylogenetic clade. We consider this filtering approach as rather conservative; it is, however, qualitative and (for now) cannot be connected to e.g. a significance level. Nonetheless, it appears reasonable to assume that random hits are considerably reduced by the multiple filtering steps at work. First of all, a gene must be present in the assembly. Second, it needs to be identified as either lost, positively selected, contracted, or expanded exclusively. Thirdly, it needs to contribute to the same mammalian phenotype as other genes which went through the same filtering steps but were categorized under a different mode of evolution, and fourth, this needs to be the case exclusively for genes from genomes of the addressed phylogenetic clade. We therefore hypothesize that if genes contributing to the same mammalian phenotype are affected by all the four modes/results of evolution (selection, contraction, expansion, loss) tested here, this strongly indicates an importance of the respective phenotype for the evolution of the respective phylogenetic clade.

By applying the described filter strategies, we did not discover any mammalian phenotype for Macroscelidea. However, we found multiple phenotypes in the clades Macroscelididae, Rhynchocyonidae, Elephantulinae, and Macroscelidinae. While the terms found within Rhynchocyonidae and Macroscelididae are more cryptic and affecting the nervous system, hematopoietic system, immune system, and locomotion, the ones found within Macroscelidinae and Elephantulinae appear (partly) to be more straightforward to interpret. Within Macroscelidinae, half of the terms detected were associated with water metabolism. Within this clade, all representatives but *P. tetradactylus* occur in arid habitats, from arid wood and bush land to desert gravel plains (Heritage, 2018; Perrin and Rathbun, 2013), suggesting a value of the underlying genes for the adaptation to arid ecosystems. For the clade Elephantulinae, a broad variety of phenotype terms was found. Besides phenotypes affecting the nervous system, the hematopoietic system, locomotion, and vision, multiple overlapping phenotypes associated with body size and skeleton morphology were discovered. Furthermore, a selection of phenotypes associated with reproduction and embryo phenotype were detected. Especially terms like "embryonic lethality between implantation and placentation" are of interest as they are in accordance with the gene ENSG00000135346, which was found to be heavily multiplied in all soft-furred sengis genomes and is associated with ectopic pregnancy and the Ovarian Hyperstimulation Syndrome. Polyovulation (also termed superovulation) has been reported for a number of soft-furred sengi species. It describes the production of more ova than developed (Birney and Baird, 1985). Usually, the number of ova shed

at ovulation is in close numerical relationship to the number of offspring produced per pregnancy in mammals, preventing wastage of the limited number of female eggs. However, some mammals, often with relatively stable litter sizes of two, shed many more ova than offspring produced. *Elephantulus myurus*, for example, sheds up to 120 eggs per cycle (Birney and Baird, 1985; Wimsatt, 1975). Although not fully understood in its evolutionary significance, polyovulation is hypothesized to be a mechanism to stabilize the number of embryos per pregnancy to two. Described by Birney and Baird (1975), this might be of evolutionary advantage for species that produce relatively large and precocial young while ensuring maneuverability e.g. to avoid predation, during pregnancy. The here detected genes might therefore play an important role in the evolutionary strategy of polyovulation in sengis and can guide future approaches for a better understanding of this phenomenon. A single gene (ENSG00000153446) was found extensively multiplied in all sengi genomes and is associated with the Jackson-Weiss syndrome which causes craniosynostosis, the premature closing of the sutures, midfacial hypoplasia, and foot anomalies (Schriml et al., 2019). Although the traits which are impacted by a specific gene might change during the course of evolution, ENSG00000153446 is a promising candidate gene, potentially involved in the shaping of the skeleton phenotype of sengis.

Among the multiplied genes of giant sengis, one (ENSG00000065809) was found to be associated with night vision difficulties in humans (ibid.). This is of significance as giant-sengis are the only exclusively diurnal sengis (Heritage, 2018; Perrin and Rathbun, 2013). Furthermore, ENSG00000151135, also multiplied within this clade, might play a role in *Rhynchocyon* skeleton morphology, as it leads to "severe congenital micromelia with shortening and distal tapering of the humeri and femora" in humans (McKusick, 2007). ENSG00000137501 was found to be multiplied in all *Elephantulus* species, potentially involved in the shaping of fur phenotype (Schriml et al., 2019).

Taken together, this study fundamentally extended the publicly available genomic resources of sengis. By applying an extensive set of tests, various aspects about the enigmatic evolution of sengis were investigated. This study demonstrates that it is possible with state-of-the-art techniques and a limited budget to do extensive genomic analysis of non-model organisms to improve our understanding of their evolutionary past.

6.6   Conclusion

The here presented study produced an unprecedented amount of genomic data
to study the evolution of sengis. Besides refining the temporal calibration of the
phylogeny of this mammalian order, we were able to exclude gene-flow and ILS as
important factors for sengi evolution. Furthermore, this study gathered multiple evi-
dence that the evolutionary past of the three species *G. rufescens*, *P. tetradactylus* and
P. rozeti notably differs from those of other sengis. By applying a recently developed,
orthology-based annotation approach (TOGA) to these non-model organisms and
linking them to mammalian phenotypes, we were able to identify phenotypic traits
which most likely played an important role for the adaptation of different phyloge-
netic entities within sengis. Furthermore, we identified genes which experienced
extensive, clade-specific multiplications and connected them to potential phenotypes
they act on. The hereby unrevealed candidate genes constitute a valuable resource for
future research on this group, as well as studies investigating evolutionary aspects of
Afrotherians or mammals as a whole.

6.7 Supplements

Please refer to the electronic version of this work to view the supplemental files, as they are in no printable format.

Chapter 7

Discussion

7.1 Summary of main results

7.1.1 New genetic and genomic resources

This thesis created a vast repository of genetic and genomic resources for nearly all sengi species. In total, it produced nuclear and mitochondrial genetic information of 79 sengi individuals from eighteen different species. The number of sengi species with available mitochondrial genomes was extended from seven to eighteen. In total, 49 complete or nearly complete new mitochondrial genomes were assembled, annotated, and uploaded to public databases, mostly extracted from museum specimens. The genomic resources of Macroscelidea were extended from a single nuclear genome to fourteen well-annotated ones, whereby two of the novel genomes were de-novo assembled. The information gathered during this study created the foundation to address a broad set of questions about the evolutionary past of Macroscelidea, which were previously impossible to target. This thesis analyzed these novel resources and investigates numerous aspects about the evolutionary past of Macroscelidea. The wealth and diversity of the data set produced here, in conjunction with previously published resources, allowed the focus on different phylogenetic levels, from the mammalian order Macroscelidea as a whole to the population genetics of *P. tetradactylus*. The data and insights collected here will, therefore, facilitate, guide, and contribute to future studies on Macroscelidea, Afrotheria, or mammals in general.

7.1.2 Macroscelidea phylogeny and evolutionary time-frame

Chapter 4 focuses on the overall phylogenetic relationships of sengi species. It constitutes the first molecular biological work that included genetic information of all recent sengi species and could, therefore, confirm the monophyly of the genus *Elephantulus*. By investigating parameter impact on the temporal calibration of phylogenetic trees, it not only clarifies the timeframe of sengi evolution but also

highlights the importance of interpreting dated phylogenetic trees in the context of the data set used to create them. It furthermore demonstrates how parameter impact on temporal calibration of phylogenies can be evaluated to draw informed conclusions and to get a sense of the node age uncertainties. Findings of this work are not only of relevance for sengi research but offer solutions on how to adequately integrate fossil calibration points from phylogenetically and biologically distant groups. This work will, therefore, be of relevance for future temporal calibration studies with limited data and a fragmented fossil record.

7.1.3   Population genetics and biogeography of *P. tetradactylus*

Chapter 5 focuses on a single species, *P. tetradactylus*. By accessing and analyzing genetic information from 55 individuals, it was able to draw conclusions about the population structure of this species, its biogeographic past, and environmental factors that shaped its evolution. This study revealed multiple, deeply divergent genetic lineages within the genus *Petrodromus*. At least one of these lineages, found in the Udzungwa mountains of Tanzania, a biodiversity hotspot and part of the Eastern Arc Mountains, was previously undescribed and therefore represents a potential candidate species. The results of this study highlight the need for a taxonomic revision of the genus *Petrodromus* in conjunction with morphological data. It furthermore identifies the dynamics of East African forests through climate cycles as a major force shaping the population structure of this, so far monotypic, soft-furred sengi genus. In contrast to previous assumptions, geological barriers like rivers and mountains are no major factors shaping the genetic structure of *Petrodromus*.

7.1.4   Evolutionary insights through comparative genomics

Chapter 6 assembles, annotates, and analyzes whole nuclear genomes of fourteen different sengi species. A data set of 8.024 one-to-one orthologous protein-coding genes enabled a confirmation and further clarification of the temporal time-frame of sengi evolution estimated in Chapter 4. Besides little morphological difference between the species, absence of impermeable geographic barriers and sympatric distribution of species dwelling in similar niches, nearly no signs of post-speciation gene flow and ILS were detected. Sengi speciation is therefore a rare example of Darwinian bifurcating evolution in genomics. An analysis of olfactory receptor proteins underlined the importance of this sensory perception for sengis, as well as revealing clade-specific differences. A reconstruction of the more recent demographic history of different sengi species and comparison to the ones of other small African mammals revealed no sengi-specific demographic patterns but a decline in effective

population size in nearly all species, starting around 1-0.8my ago. By linking genes which were either positively selected, lost, experienced gene-family contraction or expansion to mammalian phenotypic traits, this study identified multiple candidate genes of potential outstanding importance for the evolution of different sengi clades. These results were further extended by investigating clade-specific, extensively multiplied genes.

7.2   Conclusion and general perspectives

The mammalian lineage Macroscelidea and its major splits are of considerably different ages than previously assumed by multiple studies (e.g., Álvarez-Carretero et al., 2022; Douady et al., 2003; Heritage et al., 2020; Smit et al., 2011; Chapter 4; 5). Based on the most recent fossil findings and the data set created here, the MRCA of all recent sengis lived during the early Oligocene around 26-34mya (95% HPD). The split into the two families of giant sengis and soft-furred sengis has therefore unlikely been caused by enhanced competition for similar ecological niches through Afro-Laurasian faunal exchange, as it predates the major immigration waves from Eurasian mammals into Africa (Van Couvering and Delson, 2020; Chapter 4; Chapter 6). Although speculative, this split might be more likely associated with climatic and resulting ecological changes at the end of the Eocene and beginning of the Oligocene in Africa (Morley, 2000; Van Couvering and Delson, 2020). Unfortunately, our knowledge of Africas climatic, biogeographic, geological, faunal and floral past of this period is highly incomplete and oftentimes derived from only single/few sampling sites (Morley and Kingdon, 2013). However, resulting from a transition to a cooler and dryer climate, a considerable change in floral communities is documented for Africa at the end of the Eocene (Morley, 2000). Previously wide spread tropical rain forest habitats retracted towards the equator and were replaced by woodlands with an understorey of emerging grass species (Germeraad et al., 1968; Kedves, 1971; Morley, 2000; Morley and Kingdon, 2013; Salard-Cheboldaeff, 1990). Northern Africa was characterized by a mosaic of savanna, open forests and gallery forests at that time (Boureau et al., 1983). As forest distribution dynamics through climatic changes were identified as major factor shaping the biogeographic past of the genus *Petrodromus* in Chapter 5, it seems plausible that similar ecological fluctuations affected sengi evolution already earlier in time. However, to gain a clearer understanding of sengi's early evolutionary history, several knowledge gaps need to be addressed through future research. A more detailed representation of the climatic past and subsequent distribution of ecoregions through space and time in Africa would in general be beneficial. A more comprehensive and well dated fossil record is needed in order to improve our understanding of the evolution of extinct and extant sengi lineages through time. Additionally, knowledge about which external factors shape the evolution of recent sengi species dwelling in arid/desert habitats would be of interest, as this knowledge might help to identify forces which potentially influenced the evolution of similar arid adapted early sengi forms represented in the fossil record (Holroyd, 2010). These insights, in combination with the knowledge gained through the work presented here, would likely improve our understanding of early

and contemporary macroscelidean evolution. Furthermore, this knowledge should also help to clarify if the loss of macroscelidean biodiversity during the first half of the Miocene, documented in the fossil record, resulted from enhanced competition, climatic events, or both (ibid.).

Similar to the sengi crown split, the radiations of the two extant sengi families also seem unlikely to have been induced by competition through Laurasian mammals as they substantially postdate the formation of the Tauride land bridge 21mya (Pekar and Deconto, 2006; Van Couvering and Delson, 2020). The crown split of soft-furred sengis is estimated to be 8.7my old, the one of giant sengis around 500ky (Chapter 4; 6). Moreover, the here estimated node ages of the sengi phylogeny are in accordance with those suggested by Krásová and colleagues (2021), who proposed the fragmentation and reconnection of East African forest habitats through climate oscillations as the major driver for speciation events in sengis (Krásová et al., 2021).

The plausibility of this scenario was further supported by a closer examination of the biogeographic history of the monotypic genus *Petrodromus* (Chapter 5). Late Miocene aridity and the resulting fragmentation of the East African forest presumably facilitated the split between *P. tetradactylus* and its next-close relative *P. rozeti* around 5mya. While *P. rozeti* is adapted to dryer habitats, the common ancestor of *P. tetradactylus* was restricted to more humid ones. Throughout climatic dry phases and the accompanying shrinkage of forest habitats in central and eastern Africa, multiple forest refugia like the Eastern Arc Mountains and the East African coastal forests persisted. These refugia allowed forest-dwelling species to endure the arid conditions and subsequently disperse through the reconnecting forests during wetter climate phases (Bryja et al., 2017; Joordens et al., 2019; Kingdon, 2013). The findings of Chapter 5 reveal that *Petrodromus*' biogeographic past was also driven by these events. Out of the five main lineages found in *Petrodromus*, those sampled in dryer forests showed a much broader distribution. Remarkably, the distribution of one single genetic cluster spans from central Tanzania to western DRC. In addition to the inferences from biogeographic modeling, this findings strongly indicate that mountains and rivers are unlikely to pose major barriers for this taxon and do not explain its genetic structure, as previously assumed (Jennings and Rathbun, 2001, Chapter 5). The circumstance that the mentioned genetic lineage can be found on both sides of the previously assumed allopatric *Petrodromus* distribution demonstrates that this assumption most likely reflects a sampling artifact and not the actual distribution. The absence of *Petrodromus* from forest habitats north of the Congo River and in western Africa (Rathbun, 2009b) appears even more puzzling in the context of these results. However, in regard to the supposedly incorrectly assumed allopatric distribution, it might be necessary to confirm that the absence of *Petrodromus* records from north of

the Congo River is truly evidence of its absence from these areas.

Besides the implications for sengi evolution, this study (and also Chapter 4) clearly highlights the potency of museomics to investigate extant species for which sampling is difficult to conduct. The contemporary collection of a comparable comprehensive data set of *Petrodromus* fresh tissue samples can be considered close to impossible with regard to funding, responsible risk management, time, and sampling effort. It furthermore highlights the feasibility, importance, and potential of similar studies on other sengi species or genera to uncover phylogenetic patterns that can then lead to biogeographic insights and a better understanding of sengi evolution as a whole. Especially investigating which biogeographic factors shape the genetic structure of more arid-adapted sengi species could be of value, as they are largely unknown. The here discovered, yet undescribed genetic lineage from the Udzungwa Mountains of Tanzania plus the recently described giant sengi species *R. udzungwensis* from the same area highlight the importance of future sampling efforts. Besides covering the general distribution of the respective taxon, these efforts should focus on climate change refugia as well as altitudinal differences. Especially for the latter task, museum samples might be inadequate, as such information is oftentimes missing on historical specimen vouchers.

As stated in the introduction, sengis exhibit minimal morphological variation, particularly within the soft-furred sengi species, despite the fact that all species-species divergences occurred well over one million years ago (at least in soft-furred sengis; Chapter 4 and 6). The combination of the latter situation and the extensive overlap in the distributions of species occupying similar ecological niches (as illustrated in Figure 3.1e and 3.1f) might suggest that when analyzing sengi genomes, gene flow and consequently gene tree discordance should be identified. However, nearly no gene tree discordance was detected when analyzing protein coding genes as well as windowtrees of whole genomes. This makes sengis an example of classical, bifurcating speciation processes, a highly uncommon phenomenon in the age of evolutionary genomic research. A closer investigation of these findings should be conducted, including multiple *Rhynchocyon* species, as signs of mitochondrial introgression were found in this genus (Lawson et al., 2013). In this context, it could also be of value to examine the role of gene flow among populations of a single sengi species, e.g. *P. tetradactylus*. To elucidate if the nearly non existence of gene-tree discordance is a sengi-specific phenomenon in Afrotherians, or if other Afroinsectivora show similar patterns, would also be of interest. Comparative genomic approaches for tenrecs, golden moles and otter-shrews are therefore eminent. Genomic data of those three families would not only improve our understanding of their evolution, but facilitate modern scientific research on the clade Afroinsectivoria as a whole.

The low amount of gene tree discordance, which was detected regarding the three species *P. rozeti*, *P. tetradactylus*, and *G. rufescens*, points towards divergent evolutionary patterns of these species in comparison to other sengis. These findings were further supported by a low amount of shared olfactory receptor proteins compared to, for example, the genus *Macroscelides*, whose members are of similar evolutionary distance from each other as the latter three species (Chapter 6). However, the reasons for these differences are yet to be discovered. Again, population genomic data could lay the foundation to investigate the role of gene flow among the populations of each of the three species, while the comparison to other sengi genomes, could help to detect differences which separate the clade of *P. rozeti*, *P. tetradactylus*, and *G. rufescens* from other sengis.

Demographic modeling of past population sizes did not reveal a clear sengi-specific pattern, differing from other small African mammals. However, nearly all sengis experienced a decline in population size from approximately one million years ago to 20,000-10,000 years ago (Chapter 6). This timeframe was dominated by the climatic oscillations of the Pleistocene, indicating that Macroscelidea are vulnerable to climatic changes. With the climatic stabilization over the past ten thousand years, a rise in effective population size could be observed for some sengi-species. In general, sengis are not assumed to be particularly endangered (IUCN red list), although it has been mentioned that the forest-dwelling giant sengis might, in general, be more vulnerable as forests are more affected by human activities compared to more arid ecosystems (Rathbun, 2009b, sengis.org). However, the population sizes of *E. edwardii*, *P. rozeti*, *R. petersi*, and *R. cirnei* show a more or less steady decline over the past million years. The extinction risks of these species should therefore be (re-)evaluated, especially those of the two giant sengis.

The identification of multiple candidate genes, which were positively selected, experienced gene-family contraction or expansion, or were lost in specific clades, allowed the identification of phenotypic traits of potential importance for sengis. One particular discovery worth highlighting is the identification of a gene that may be associated with the polyovulation syndrome in soft-furred sengis. The polyovulation syndrome is a yet poorly understood reproductive phenomenon in some mammal species, often with a relatively stable litter size of two. Females of these species shed vastly more ova (> 140 in *M. myurus*) than being developed (Chapter 6). Generally, the majority of phenotypic traits identified in this study were related to reproduction, physiology, or were associated with the immune and nervous systems. These less displayed characteristics could suggest that functional biological distinctions between sengi species or clades exist but may be more cryptic than previously anticipated and therefore overseen in approaches that focused on behavior and morphology. A closer

investigation of these traits in future studies might help to identify features that facilitate the biological differences between sengi species in terms of habitat usage and occurrence in different climatic zones, besides their little displayed morphological differences.

In summary, this thesis produced an amount of genetic and genomic resources unparalleled for any Afroinsectivora family so far. The approach of utilizing museum samples to investigate the molecular evolution of species for which contemporary sampling is difficult to conduct has proven its feasibility and investigative power. Analyzing the here produced data has revealed multiple insights about the evolutionary past of Macroscelidea, from basic taxonomic questions to clade-specific, functional implications of genes with potential adaptive value. These insights and the revealed knowledge gaps of our understanding of the evolution of this mammalian order will provide valuable resources and guidance for future research approaches. The here presented research can serve as a blueprint for future investigations on other poorly studied Afrotherian clades and species, while the provided genomic resources will then enable cross-family comparisons for a better understanding of Afrotheria evolution as a whole.

List of Figures

Bibliography

Agne, Stefanie, Merrick Ekins, Adrian Galitz, Michael Hofreiter, Michaela Preick, Nicolas Straube, Gert Wörheide, and Dirk Erpenbeck (2022a). *Keratose sponge MuseOMICS: setting reference points in dictyoceratid demosponge phylogeny*. DOI: 10.11646/zootaxa.5195.3.9. URL: https://pubmed.ncbi.nlm.nih.gov/37045286/.

Agne, Stefanie, Gavin J P Naylor, Michaela Preick, Lei Yang, Ralf Thiel, Simon Weigmann, Johanna L.A. Paijmans, Axel Barlow, Michael Hofreiter, Nicolas Straube, and Kieren James Mitchell (2022b). "Taxonomic Identification of Two Poorly Known Lantern Shark Species Based on Mitochondrial DNA From Wet-Collection Paratypes". In: *Frontiers in Ecology and Evolution* 10.6, pp. 1–10. DOI: 10.3389/fevo.2022.910009.

Agne, Stefanie, Michaela Preick, Nicolas Straube, and Michael Hofreiter (2022c). "Simultaneous Barcode Sequencing of Diverse Museum Collection Specimens Using a Mixed RNA Bait Set". In: *Frontiers in Ecology and Evolution* 10.6, pp. 1–5. ISSN: 2296701X. DOI: 10.3389/fevo.2022.909846.

Aleksander, Suzi A, James Balhoff, Seth Carbon, J Michael Cherry, Harold J Drabkin, Dustin Ebert, Marc Feuermann, Pascale Gaudet, Nomi L Harris, David P Hill, Raymond Lee, Huaiyu Mi, Sierra Moxon, Christopher J Mungall, Anushya Muruganugan, Tremayne Mushayahama, Paul W Sternberg, Paul D Thomas, Kimberly Van Auken, Jolene Ramsey, Deborah A Siegele, Rex L Chisholm, Petra Fey, Maria Cristina Aspromonte, Maria Victoria Nugnes, Federica Quaglia, Silvio Tosatto, Michelle Giglio, Suvarna Nadendla, Giulia Antonazzo, Helen Attrill, Gil dos Santos, Steven Marygold, Victor Strelets, Christopher J Tabone, Jim Thurmond, Pinglei Zhou, Saadullah H Ahmed, Praoparn Asanitthong, Diana Luna Buitrago, Meltem N Erdol, Matthew C Gage, Mohamed Ali Kadhum, Kan Yan Chloe Li, Miao Long, Aleksandra Michalak, Angeline Pesala, Armalya Pritazahra, Shirin C C Saverimuttu, Renzhi Su, Kate E Thurlow, Ruth C Lovering, Colin Logie, Snezhana Oliferenko, Judith Blake, Karen Christie, Lori Corbani, Mary E Dolan, Harold J Drabkin, David P Hill, Li Ni, Dmitry Sitnikov, Cynthia Smith, Alayne Cuzick, James Seager, Laurel Cooper, Justin Elser, Pankaj Jaiswal, Parul Gupta, Pankaj Jaiswal, Sushma Naithani, Manuel Lera-Ramirez, Kim Rutherford, Valerie Wood, Jeffrey L De Pons, Melinda R Dwinell, G Thomas Hayman, Mary L Kaldunski, Anne E Kwitek, Stanley J F Laulederkind, Marek A Tutaj, Mahima Vedi, Shur-Jen

Wang, Peter D'Eustachio, Lucila Aimo, Kristian Axelsen, Alan Bridge, and Hyka-Nouspikel (2023). "The Gene Ontology knowledgebase in 2023". In: *GENETICS* 224.1. Ed. by A Baryshnikova. ISSN: 1943-2631. DOI: 10.1093/genetics/iyad031. URL: https://academic.oup.com/genetics/article/doi/10.1093/genetics/iyad031/7068118.

Allio, Rémi, Alex Schomaker-Bastos, Jonathan Romiguier, Francisco Prosdocimi, Benoit Nabholz, and Frédéric Delsuc (2020). "MitoFinder: Efficient automated large-scale extraction of mitogenomic data in target enrichment phylogenomics". In: *Molecular Ecology Resources* 20.4, pp. 892–905. ISSN: 17550998. DOI: 10.1111/1755-0998.13160.

Álvarez-Carretero, Sandra, Asif U. Tamuri, Matteo Battini, Fabrícia F. Nascimento, Emily Carlisle, Robert J. Asher, Ziheng Yang, Philip C.J. Donoghue, and Mario dos Reis (2022). "A species-level timeline of mammal evolution integrating phylogenomic data". In: *Nature* 602.7896, pp. 263–267. ISSN: 14764687. DOI: 10.1038/s41586-021-04341-1. URL: https://doi.org/10.1038/s41586-021-04341-1.

Amin, Rajan, Tim Wacher, Bernard Ogwoka, Cedric Khayale, Linus Kariuki, and Bernard Agwanda (2021). "Habitat use of the endangered golden-rumped sengi Rhynchocyon chrysopygus". In: *African Journal of Ecology* 59.1, pp. 108–116. ISSN: 0141-6707. DOI: 10.1111/aje.12804. URL: https://onlinelibrary.wiley.com/doi/10.1111/aje.12804.

Amrine-Madsen, Heather, Klaus Peter Koepfli, Robert K. Wayne, and Mark S. Springer (2003). "A new phylogenetic marker, apolipoprotein B, provides compelling evidence for eutherian relationships". In: *Molecular Phylogenetics and Evolution*. Vol. 28. 2. Academic Press Inc., pp. 225–240. DOI: 10.1016/S1055-7903(03)00118-0. URL: https://pubmed.ncbi.nlm.nih.gov/12878460/.

Andanje, Samuel, Bernard Risky Agwanda, Grace W. Ngaruiya, Rajan Amin, and Galen B. Rathbun (2010). "Sengi (Elephant-Shrew) Observations from Northern Coastal Kenya". In: *Journal of East African Natural History* 99.1, pp. 1–8. ISSN: 0012-8317. DOI: 10.2982/028.099.0101.

anders, Larsson (2014). "AliView: A fast and lightweight alignment viewer and editor for large datasets". In: *Bioinformatics* 30.22, pp. 3276–3278. ISSN: 14602059. DOI: 10.1093/bioinformatics/btu531. URL: https://pubmed.ncbi.nlm.nih.gov/25095880/.

andrew, Rambaut, Alexei J. Drummond, Dong Xie, Guy Baele, and Marc A. Suchard (2018). "Posterior summarization in Bayesian phylogenetics using Tracer 1.7". In: *Systematic Biology* 67.5, pp. 901–904. ISSN: 1076836X. DOI: 10.1093/sysbio/syy032.

Árnason, Úlfur, Fritjof Lammers, Vikas Kumar, Maria A. Nilsson, and Axel Janke (2018). "Whole-genome sequencing of the blue whale and other rorquals finds signatures for introgressive gene flow". In: *Science Advances* 4.4. ISSN: 23752548. DOI: 10.1126/sciadv.aap9873. URL: https://www.science.org.

Ashburner, Michael, Catherine A. Ball, Judith A. Blake, David Botstein, Heather Butler, J. Michael Cherry, Allan P. Davis, Kara Dolinski, Selina S. Dwight, Janan T. Eppig, Midori A. Harris, David P. Hill, Laurie Issel-Tarver, Kasarskis andrew, Suzanna Lewis, John C. Matese, Joel E. Richardson, Martin Ringwald, Gerald M. Rubin, and Gavin Sherlock (2000). *Gene ontology: Tool for the unification of biology*. DOI: 10.1038/75556. URL: http://www.flybase.bio.indiana.eduhttp://fruitfly.bdgp.berkeley.eduhttp://genome-www.stanford.eduhttp://www.informatics.jax.org.

Asher, Robert J. and Thomas Lehmann (2008). "Dental eruption in afrotherian mammals". In: *BMC Biology* 6.1, p. 14. ISSN: 17417007. DOI: 10.1186/1741-7007-6-14. URL: https://bmcbiol.biomedcentral.com/articles/10.1186/1741-7007-6-14.

Balmori-de la Puente, Alfonso, Jacint Ventura, Marcos Miñarro, Aitor Somoano, Jody Hey, and Jose Castresana (2022). "Divergence time estimation using ddRAD data and an isolation-with-migration model applied to water vole populations of Arvicola". In: *Scientific Reports* 12.1, pp. 1–11. ISSN: 20452322. DOI: 10.1038/s41598-022-07877-y. URL: https://doi.org/10.1038/s41598-022-07877-y.

Barlow, Axel, James A. Cahill, Stefanie Hartmann, Christoph Theunert, Georgios Xenikoudakis, Gloria G. Fortes, Johanna L.A. Paijmans, Gernot Rabeder, Christine Frischauf, Aurora Grandal-d'Anglade, Ana García-Vázquez, Marine Murtskhval-adze, Urmas Saarma, Peeter Anijalg, Tomaž Skrbinšek, Giorgio Bertorelle, Boris Gasparian, Guy Bar-Oz, Ron Pinhasi, Montgomery Slatkin, Love Dalén, Beth Shapiro, and Michael Hofreiter (2018). "Partial genomic survival of cave bears in living brown bears". In: *Nature Ecology and Evolution* 2.10, pp. 1563–1570. ISSN: 2397334X. DOI: 10.1038/s41559-018-0654-8.

Benton, Michael J. and Francisco J. Ayala (2003). *Dating the tree of life*. DOI: 10.1126/science.1077795. URL: https://www.science.org.

Benton, Michael J., Philip C.J. Donoghue, Robert J. Asher, Matt Friedman, Thomas J. Near, and Jakob Vinther (2015). "Constraints on the timescale of animal evolutionary history". In: *Palaeontologia Electronica* 18.1, pp. 1–107. ISSN: 10948074. DOI: 10.26879/424.

Bibi, Faysal (2013). "A multi-calibrated mitochondrial phylogeny of extant Bovidae (Artiodactyla, Ruminantia) and the importance of the fossil record to systematics".

In: *BMC Evolutionary Biology* 13.1, p. 1. ISSN: 14712148. DOI: 10.1186/1471-2148-13-166. URL: BMCEvolutionaryBiology.

Birney, Elmer C. and Donna Day Baird (1985). "Why Do Some Mammals Polyovulate to Produce a Litter of Two?" In: *The American Naturalist* 126.1, pp. 136–140. ISSN: 0003-0147. DOI: 10.1086/284403.

Blake, Judith A., Richard Baldarelli, James A. Kadin, Joel E. Richardson, Cynthia L. Smith, and Carol J. Bult (2021). "Mouse Genome Database (MGD): Knowledgebase for mouse-human comparative biology". In: *Nucleic Acids Research* 49.D1, pp. D981–D987. ISSN: 13624962. DOI: 10.1093/nar/gkaa1083. URL: https://pubmed.ncbi.nlm.nih.gov/33231642/.

Bobe, R. (2006). "The evolution of arid ecosystems in eastern Africa". In: *Journal of Arid Environments* 66.3 SPEC. ISS. Pp. 564–584. ISSN: 1095922X. DOI: 10.1016/j.jaridenv.2006.01.010.

Borowiec, Marek L. (2016). "AMAS: a fast tool for alignment manipulation and computing of summary statistics". In: *PeerJ* 4.1, e1660. ISSN: 2167-8359. DOI: 10.7717/peerj.1660. URL: https://peerj.com/articles/1660.

Bouckaert, Remco, Joseph Heled, Denise Kühnert, Tim Vaughan, Chieh Hsi Wu, Dong Xie, Marc A. Suchard, Rambaut andrew, and Alexei J. Drummond (2014). "BEAST 2: A Software Platform for Bayesian Evolutionary Analysis". In: *PLoS Computational Biology* 10.4, pp. 1–6. ISSN: 15537358. DOI: 10.1371/journal.pcbi.1003537.

Boureau, E., M. Cheboldaeff-Salard, J. C. Koeniguer, and P. Louvet (1983). "Evolution des flores et de la végétation Tertiaires en Afrique, au nord de PEquateur". In: *Bothalia* 14.3/4, pp. 355–367. ISSN: 0006-8241. DOI: 10.4102/abc.v14i3/4.1180. URL: https://journals.abcjournal.aosis.co.za/index.php/abc/article/view/1180.

Bromham, Lindell (2011). "The genome as a life-history character: Why rate of molecular evolution varies between mammal species". In: *Philosophical Transactions of the Royal Society B: Biological Sciences* 366.1577, pp. 2503–2513. ISSN: 09628436. DOI: 10.1098/rstb.2011.0014.

Bronner, Gary N, Samantha Mynhardt, Nigel C Bennett, Lientjie Cohen, Nick Crumpton, Michael Hofreiter, Patrick Arnold, and Robert J Asher (2023). "Phylogenetic history of golden moles and tenrecs (Mammalia: Afrotheria)". In: *Zoological Journal of the Linnean Society*. ISSN: 0024-4082. DOI: 10.1093/zoolinnean/zlad121. URL: https://academic.oup.com/zoolinnean/advance-article/doi/10.1093/zoolinnean/zlad121/7285756.

Bryja, Josef, Radim Šumbera, Julian C. Kerbis Peterhans, Tatiana Aghová, Anna Bryjová, Ondřej Mikula, Violaine Nicolas, Christiane Denys, and Erik Verheyen

(2017). "Evolutionary history of the thicket rats (genus Grammomys) mirrors the evolution of African forests since late Miocene". In: *Journal of Biogeography* 44.1, pp. 182–194. ISSN: 13652699. DOI: 10.1111/jbi.12890.

Burgess, N. D., G. P. Clarke, and W. A. Rodgers (1998). "Coastal forests of eastern Africa: status, endemism patterns and their potential causes". In: *Biological Journal of the Linnean Society* 64.3, pp. 337–367. ISSN: 00244066. DOI: 10.1111/j.1095-8312.1998.tb00337.x. URL: https://academic.oup.com/biolinnean/article-lookup/doi/10.1111/j.1095-8312.1998.tb00337.x.

Butler, P. M. (1995). "Fossil Macroscelidea". In: *Mammal Review* 25.1-2, pp. 3–14. ISSN: 13652907. DOI: 10.1111/j.1365-2907.1995.tb00432.x.

Carlen, Elizabeth J., Galen B. Rathbun, Link E. Olson, Christopher A. Sabuni, William T. Stanley, and John P. Dumbacher (2017). "Reconstructing the molecular phylogeny of giant sengis (Macroscelidea; Macroscelididae; Rhynchocyon)". In: *Molecular Phylogenetics and Evolution* 113, pp. 150–160. ISSN: 10959513. DOI: 10.1016/j.ympev.2017.05.012. URL: http://dx.doi.org/10.1016/j.ympev.2017.05.012.

Castañeda, Isla S., Stefan Mulitza, Enno Schefuß, Raquel A.Lopes Dos Santos, Jaap S.Sinninghe Damsté, and Stefan Schouten (2009). "Wet phases in the Sahara/Sahel region and human migration patterns in North Africa". In: *Proceedings of the National Academy of Sciences of the United States of America* 106.48, pp. 20159–20163. ISSN: 00278424. DOI: 10.1073/pnas.0905771106. URL: www.pnas.org/cgi/content/full/.

Chevreux, Bastien, T Wetter, and S Suhai (1999). "Genome Sequence Assembly Using Trace Signals and Additional Sequence Information". In: *Computer Science and Biology: Proceedings of the German Conference on Bioinformatics (GCB) '99, GCB, Hannover, Germany.* 1995, pp. 45–56.

Corbet, G. B. and J. Hanks (1968). "A revision of the elephant-shrews, Family Macroscelididae". In: *Bulletin of the British Museum (Natural History) Zoology* 16, pp. 47–111. ISSN: 0007-1498. URL: http://biostor.org/reference/97993.

Corbet, G. B. and B. R. Neal (1965). "The taxonomy of the elephant shrews of the genusPetrodromus,with particular reference to the East African coast". In: *Revue de Zoologie et de Botanique Africaine* 71, pp. 49–78.

Couvreur, Thomas L.P., Lars W. Chatrou, Marc S.M. Sosef, and James E. Richardson (2008). "Molecular phylogenetics reveal multiple tertiary vicariance origins of the African rain forest trees". In: *BMC Biology* 6.1, pp. 1–10. ISSN: 17417007. DOI: 10.1186/1741-7007-6-54. URL: http://www.biomedcentral.com/1741-7007/6/54.

Couvreur, Thomas L.P., Gilles Dauby, Anne Blach-Overgaard, Vincent Deblauwe, Steven Dessein, Vincent Droissart, Oliver J. Hardy, David J. Harris, Steven B.

Janssens, Alexandra C. Ley, Barbara A. Mackinder, Bonaventure Sonké, Marc S.M. Sosef, Tariq Stévart, Jens-Christian Svenning, Jan J. Wieringa, Adama Faye, Alain D. Missoup, Krystal A. Tolley, Violaine Nicolas, Stéphan Ntie, Frédiéric Fluteau, Cécile Robin, Francois Guillocheau, Doris Barboni, and Pierre Sepulchre (2021). "Tectonics, climate and the diversification of the tropical African terrestrial flora and fauna". In: *Biological Reviews* 96.1, pp. 16–51. ISSN: 1464-7931. DOI: 10.1111/brv.12644. URL: https://onlinelibrary.wiley.com/doi/10.1111/brv.12644.

Cunningham, Fiona, James E. Allen, Jamie Allen, Jorge Alvarez-Jarreta, M. Ridwan Amode, Irina M. Armean, Olanrewaju Austine-Orimoloye, Azov andrey G., If Barnes, Ruth Bennett, Berry andrew, Jyothish Bhai, Alexandra Bignell, Konstantinos Billis, Sanjay Boddu, Lucy Brooks, Mehrnaz Charkhchi, Carla Cummins, Luca Da Rin Fioretto, Claire Davidson, Kamalkumar Dodiya, Sarah Donaldson, Bilal El Houdaigui, Tamara El Naboulsi, Reham Fatima, Carlos Garcia Giron, Thiago Genez, Jose Gonzalez Martinez, Cristina Guijarro-Clarke, Arthur Gymer, Matthew Hardy, Zoe Hollis, Thibaut Hourlier, Toby Hunt, Thomas Juettemann, Vinay Kaikala, Mike Kay, Ilias Lavidas, Tuan Le, Diana Lemos, Jose Carlos Marugán, Shamika Mohanan, Aleena Mushtaq, Marc Naven, Denye N. Ogeh, Anne Parker, Parton andrew, Malcolm Perry, Ivana Pilizota, Irina Prosovetskaia, Manoj Pandian Sakthivel, Ahamed Imran Abdul Salam, Bianca M. Schmitt, Helen Schuilenburg, Dan Sheppard, Jose G. Perez-Silva, William Stark, Emily Steed, Kyösti Sutinen, Ranjit Sukumaran, Dulika Sumathipala, Marie Marthe Suner, Michal Szpak, Anja Thormann, Francesca Floriana Tricomi, David Urbina-Gómez, Veidenberg andres, Thomas A. Walsh, Brandon Walts, Natalie Willhoft, Winterbottom andrea, Elizabeth Wass, Marc Chakiachvili, Bethany Flint, Adam Frankish, Stefano Giorgetti, Leanne Haggerty, Sarah E. Hunt, Garth R. Iisley, Jane E. Loveland, Fergal J. Martin, Benjamin Moore, Jonathan M. Mudge, Matthieu Muffato, Emily Perry, Magali Ruffier, John Tate, David Thybert, Stephen J. Trevanion, Sarah Dyer, Peter W. Harrison, Kevin L. Howe, Yates andrew D., Daniel R. Zerbino, and Paul Flicek (2022). "Ensembl 2022". In: *Nucleic Acids Research* 50.D1, pp. D988–D995. ISSN: 13624962. DOI: 10.1093/nar/gkab1049. URL: https://doi.org/10.1093/nar/gkab1049.

Dabney, Jesse, Michael Knapp, Isabelle Glocke, Marie Theres Gansauge, Antje Weihmann, Birgit Nickel, Cristina Valdiosera, Nuria García, Svante Pääbo, Juan Luis Arsuaga, and Matthias Meyer (2013). "Complete mitochondrial genome sequence of a Middle Pleistocene cave bear reconstructed from ultrashort DNA fragments". In: *Proceedings of the National Academy of Sciences of the United States of America* 110.39, pp. 15758–15763. ISSN: 00278424. DOI: 10.1073/pnas.1314445110.

Davenport, Tim R.B., William T. Stanley, Eric J. Sargis, Daniela W. De Luca, Noah E. Mpunga, Sophy J. Machaga, and Link E. Olson (2006). "A new genus of African

monkey, Rungwecebus: Morphology, ecology and molecular phylogenetics". In: *Science* 312.5778, pp. 1378–1381. ISSN: 00368075. DOI: 10.1126/science.1125631. URL: www.sciencemag.org/cgi/content/full/312/5778/1375/DC1.

De Abreu-Jr, Edson Fiedler, Edson Fiedler De Abreu-Jr, Silvia E. Pavan, Silvia E. Pavan, Mirian T.N. Tsuchiya, Mirian T.N. Tsuchiya, Don E. Wilson, Alexandre R. Percequillo, Jesús E. Maldonado, and Jesús E. Maldonado (2020). "Museomics of tree squirrels: A dense taxon sampling of mitogenomes reveals hidden diversity, phenotypic convergence and the need of a taxonomic overhaul". In: *BMC Evolutionary Biology* 20.1, p. 77. ISSN: 14712148. DOI: 10.1186/s12862-020-01639-y. URL: https://bmcevolbiol.biomedcentral.com/articles/10.1186/s12862-020-01639-y.

De Jong, Wilfried W., Anneke Zweers, and Morris Goodman (1981). "Relationship of aardvark to elephants, hyraxes and sea cows from α-crystallin sequences". In: *Nature* 292.5823, pp. 538–540. ISSN: 00280836. DOI: 10.1038/292538a0. URL: https://pubmed.ncbi.nlm.nih.gov/7254349/.

Dierckxsens, Nicolas, Patrick Mardulyn, and Guillaume Smits (2017). "NOVOPlasty: De novo assembly of organelle genomes from whole genome data". In: *Nucleic Acids Research* 45.4, p. 18. ISSN: 13624962. DOI: 10.1093/nar/gkw955. URL: http://www.ebi.ac.uk.

Dinerstein, Eric, David Olson, Anup Joshi, Carly Vynne, Neil D. Burgess, Eric Wikramanayake, Nathan Hahn, Suzanne Palminteri, Prashant Hedao, Reed Noss, Matt Hansen, Harvey Locke, Erle C Ellis, Benjamin Jones, Charles Victor Barber, Randy Hayes, Cyril Kormos, Vance Martin, Eileen Crist, Wes Sechrest, Lori Price, Jonathan E. M. Baillie, Don Weeden, Kierán Suckling, Crystal Davis, Nigel Sizer, Rebecca Moore, David Thau, Tanya Birch, Peter Potapov, Svetlana Turubanova, Alexandra Tyukavina, Nadia de Souza, Lilian Pintea, José C. Brito, Othman A. Llewellyn, Anthony G. Miller, Annette Patzelt, Shahina A. Ghazanfar, Jonathan Timberlake, Heinz Klöser, Yara Shennan-Farpón, Roeland Kindt, Jens-Peter Barnekow Lillesø, Paulo van Breugel, Lars Graudal, Maianna Voge, Khalaf F. Al-Shammari, and Muhammad Saleem (2017). "An Ecoregion-Based Approach to Protecting Half the Terrestrial Realm". In: *BioScience* 67.6, pp. 534–545. ISSN: 0006-3568. DOI: 10.1093/biosci/bix014. URL: https://academic.oup.com/bioscience/article-lookup/doi/10.1093/biosci/bix014.

Dinesen, Lars, Thomas Lehmberg, Marcel C. Rahner, and Jon Fjeldså (2001). "Conservation priorities for the forests of the Udzungwa Mountains, Tanzania, based on primates, duikers and birds". In: *Biological Conservation* 99.2, pp. 223–236. ISSN: 00063207. DOI: 10.1016/S0006-3207(00)00218-4.

Donoghue, Philip C.J. and Michael J. Benton (2007). *Rocks and clocks: calibrating the Tree of Life using fossils and molecules*. DOI: 10.1016/j.tree.2007.05.005.

Dornburg, Alex, Jeffrey P. Townsend, Matt Friedman, and Thomas J. Near (2014). "Phylogenetic informativeness reconciles ray-finned fish molecular divergence times". In: *BMC Evolutionary Biology* 14.1, p. 169. ISSN: 14712148. DOI: 10.1186/s12862-014-0169-0. URL: http://bmcevolbiol.biomedcentral.com/articles/10.1186/s12862-014-0169-0.

Douady, Christophe J. (2001). "Molecular Phylogenetics of the insectivoria". In: *PhD Dissertation, The Queen's University of Belfast*.

Douady, Christophe J., François Catzeflis, Jaishree Raman, Mark S. Springer, and Michael J. Stanhope (2003). "The Sahara as a vicariant agent and the role of miocene climatic events, in the diversification of the mammalian order Macroscelidea (elephant shrews)". In: *Proceedings of the National Academy of Sciences of the United States of America* 100.14, pp. 8325–8330. ISSN: 00278424. DOI: 10.1073/pnas.0832467100.

Douglas, Jordan, Rong Zhang, and Remco Bouckaert (2021). "Adaptive dating and fast proposals: Revisiting the phylogenetic relaxed clock model". In: *PLoS Computational Biology* 17.2, e1008322. ISSN: 15537358. DOI: 10.1371/JOURNAL.PCBI.1008322. URL: https://doi.org/10.1371/journal.pcbi.1008322.

Drummond, Alexei J. and Rambaut andrew (2007). "BEAST: Bayesian evolutionary analysis by sampling trees". In: *BMC Evolutionary Biology* 7.1, pp. 1–8. ISSN: 14712148. DOI: 10.1186/1471-2148-7-214.

Drummond, Alexei J. and Marc A. Suchard (2010). "Bayesian random local clocks, or one rate to rule them all". In: *BMC Biology* 8.1, p. 114. ISSN: 17417007. DOI: 10.1186/1741-7007-8-114. URL: https://bmcbiol.biomedcentral.com/articles/10.1186/1741-7007-8-114.

Dumbacher, John P., Elizabeth J. Carlen, and Galen B. Rathbun (2016). "Petrosaltator gen. Nov., a new genus replacement for the North African sengi Elephantulus rozeti (Macroscelidea; Macroscelididae)". In: *Zootaxa* 4136.3, pp. 567–579. ISSN: 11755334. DOI: 10.11646/zootaxa.4136.3.8.

Dumbacher, John P., Galen B. Rathbun, Hanneline A. Smit, and Seth J. Eiseb (2012). "Phylogeny and taxonomy of the round-eared sengis or elephant-shrews, genus Macroscelides (Mammalia, Afrotheria, Macroscelidea)". In: *PLoS ONE* 7.3. ISSN: 19326203. DOI: 10.1371/journal.pone.0032410.

Fitzgibbon, Clare D (1995). "Comparative ecology of two elephant-shrew species in a Kenyan coastal forest". In: *Mammal Review* 25, pp. 19–30.

Foley, Nicole M., Mark S. Springer, and Emma C. Teeling (2016). *Mammal madness: Is the mammal tree of life not yet resolved?* DOI: 10.1098/rstb.2015.0140.

URL: http://dx.doi.org/10.1098/rstb.2015.0140orviahttp://rstb.royalsocietypublishing.org..

Fontsere, Claudia, Marc de Manuel, Tomas Marques-Bonet, and Martin Kuhlwilm (2019). "Admixture in Mammals and How to Understand Its Functional Implications: On the Abundance of Gene Flow in Mammalian Species, Its Impact on the Genome and Roads into a Functional Understanding". In: *BioEssays* 41.12, p. 1900123. ISSN: 15211878. DOI: 10.1002/bies.201900123. URL: https://onlinelibrary.wiley.com/doi/full/10.1002/bies.201900123https://onlinelibrary.wiley.com/doi/abs/10.1002/bies.201900123https://onlinelibrary.wiley.com/doi/10.1002/bies.201900123.

Gannamani, Rahul, Sterre van der Veen, Martje van Egmond, Tom J. de Koning, and Marina A.J. Tijssen (2021). "Challenges in Clinicogenetic Correlations: One Phenotype – Many Genes". In: *Movement Disorders Clinical Practice* 8.3, pp. 311–321. ISSN: 2330-1619. DOI: 10.1002/mdc3.13163. URL: https://movementdisorders.onlinelibrary.wiley.com/doi/10.1002/mdc3.13163.

Gansauge, Marie Theres, Tobias Gerber, Isabelle Glocke, Petra Korlević, Laurin Lippik, Sarah Nagel, Lara Maria Riehl, Anna Schmidt, and Matthias Meyer (2017). "Single-stranded DNA library preparation from highly degraded DNA using T4 DNA ligase". In: *Nucleic Acids Research* 45.10. ISSN: 13624962. DOI: 10.1093/nar/gkx033.

Gansauge, Marie Theres and Matthias Meyer (2013). "Single-stranded DNA library preparation for the sequencing of ancient or damaged DNA". In: *Nature Protocols* 8.4, pp. 737–748. ISSN: 17542189. DOI: 10.1038/nprot.2013.038.

Germeraad, J. H., C. A. Hopping, and J. Muller (1968). "Palynology of tertiary sediments from tropical areas". In: *Review of Palaeobotany and Palynology* 6.3-4, pp. 189–348. ISSN: 00346667. DOI: 10.1016/0034-6667(68)90051-1.

Gheerbrant, E., J. Sudre, S. Sen, C. Abrial, B. Marandat, B. Sigé, and M. Vianey-liaud (1998). "Nouvelles données sur les mammifères du Thanétien et de l'Ypresien du Bassin d'Ouarzazate (Maroc) et leur contexte stratigraphique". In: *undefined*.

Gheerbrant, Emmanuel (2009). "Paleocene emergence of elephant relatives and the rapid radiation of African ungulates". In: *Proceedings of the National Academy of Sciences of the United States of America* 106.26, pp. 10717–10721. ISSN: 00278424. DOI: 10.1073/pnas.0900251106.

Gheerbrant, Emmanuel, Mbarek Amaghzaz, Baadi Bouya, Florent Goussard, and Charlène Letenneur (2014). "Ocepeia (middle Paleocene of Morocco): The oldest skull of an afrotherian mammal". In: *PLoS ONE* 9.2. ISSN: 19326203. DOI: 10.1371/journal.pone.0089739.

Gheerbrant, Emmanuel, Jean Sudre, Mohamed Iarochene, and Abdelkader Moumni (2001). "First ascertained african "condylarth" mammals (primitive ungulates: Cf. Bulbulodentata and cf. phenacodonta) from the earliest ypresian of the ouled abdoun basin, Morocco". In: *Journal of Vertebrate Paleontology* 21.1, pp. 107–118. ISSN: 19372809. DOI: 10.1671/0272-4634(2001)021[0107:FAACMP]2.0.CO;2.

Graur, Dan and William Martin (2004). "Reading the entrails of chickens: Molecular timescales of evolution and the illusion of precision". In: *Trends in Genetics* 20.2, pp. 80–86. ISSN: 01689525. DOI: 10.1016/j.tig.2003.12.003.

Grubb, P., O. Sandrock, O. Kullmer, T. M. Kaiser, and F. Schrenk (1999). *Relationships between eastern and southern African mammal faunas. African biogeography, climate change and human evolution*. New York: Oxford University Press, pp. 107–115.

Guindon, Stéphane (2018). "Accounting for Calibration Uncertainty: Bayesian Molecular Dating as a "Doubly Intractable" Problem". In: *Systematic Biology* 67.4. Ed. by David Bryant, pp. 651–661. ISSN: 1063-5157. DOI: 10.1093/sysbio/syy003. URL: https://academic.oup.com/sysbio/article/67/4/651/4827614.

Hagemann, Justus, Michael Hofreiter, Faysal Bibi, Patricia Holroyd, and Patrick Arnold (2023). "Is it inappropriate to ask for your age? Evaluating Parameter Impact on Tree Dating in a Challenging Clade (Macroscelidea)". In: *Molecular Phylogenetics and Evolution* 183, p. 107756. ISSN: 10557903. DOI: 10.1016/j.ympev.2023.107756.

Hahn, Christoph, Lutz Bachmann, and Bastien Chevreux (2013). "Reconstructing mitochondrial genomes directly from genomic next-generation sequencing reads - A baiting and iterative mapping approach". In: *Nucleic Acids Research* 41.13. ISSN: 03051048. DOI: 10.1093/nar/gkt371.

Hahn, M W (2018). *Molecular Population Genetics*. Sinauer Series. Oxford University Press, "Chapter 9, Demographic History", Molecular Popula. ISBN: 9780878939657.

Heath, Tracy A., John P. Huelsenbeck, and Tanja Stadler (2014). "The fossilized birth-death process for coherent calibration of divergence-time estimates". In: *Proceedings of the National Academy of Sciences of the United States of America* 111.29, E2957–E2966. ISSN: 10916490. DOI: 10.1073/pnas.1319091111. URL: https://www.pnas.org/doi/abs/10.1073/pnas.1319091111.

Heine, Bernd and Wilhelm J. G. Möhling (1980). *Language and Dialect Atlas of Kenya*. Vol. 1. Dietrich Reimer Verlag, pp. 17–22. DOI: 10.1075/lplp.7.1.14lai.

Hempel, Elisabeth, Michael V. Westbury, José H. Grau, Alexandra Trinks, Johanna L.A. Paijmans, Sergei Kliver, Axel Barlow, Frieder Mayer, Johannes Müller, Lei Chen,

Klaus Peter Koepfli, Michael Hofreiter, and Faysal Bibi (2021). "Diversity and paleodemography of the addax (Addax nasomaculatus), a saharan antelope on the verge of extinction". In: *Genes* 12.8. ISSN: 20734425. DOI: 10.3390/genes12081236.

Heritage, Steven (2018). "Family Macroscelididae (Sengis)". In: *Handbook of the mammals of the world, volume 8, insectivores, sloths and colugos.* Pp. 206–234. ISBN: 9788416728084.

Heritage, Steven, Houssein Rayaleh, Djama G Awaleh, and Galen B. Rathbun (2020). "New records of a lost species and a geographic range expansion for sengis in the Horn of Africa". In: pp. 1–38. DOI: 10.7717/peerj.9652.

Ho, Simon Y. W. and Matthew J. Phillips (2009). "Accounting for Calibration Uncertainty in Phylogenetic Estimation of Evolutionary Divergence Times". In: *Systematic Biology* 58.3, pp. 367–380. ISSN: 1076-836X. DOI: 10.1093/sysbio/syp035. URL: https://academic.oup.com/sysbio/article/58/3/367/1613552.

Hofreiter, Michael (2012). "Nondestructive DNA extraction from museum specimens". In: *Methods in Molecular Biology* 840, pp. 93–100. ISSN: 10643745. DOI: 10.1007/978-1-61779-516-9_13. URL: https://link.springer.com/protocol/10.1007/978-1-61779-516-9{_}13.

Hofreiter, Michael, David Serre, Hendrik N Poinar, Melanie Kuch, and Svante Pääbo (2001). "Hofreiter_Ancient_NatRevGen_2001_1556040". In: *Nature.Com* 2.May, pp. 3–9.

Hollister, N. (1917). "Descriptions of a new genus and eight new species of African mammals". In: *Smithsonian miscellaneous collections 66*, pp. 1–8.

Holroyd, Patricia (2010). "Macroscelidea". In: *Cenozoic mammals of Africa*. Ed. by Lars Werdelin and William Joseph Sanders. Berkeley: University of California Press. Chap. 8, pp. 89–98. DOI: 10.5860/choice.48-2092. URL: https://www.worldcat.org/title/cenozoic-mammals-of-africa/oclc/449860225?referer=di{\&}ht=edition.

Horn, Susanne (2012). "Target enrichment via DNA hybridization capture". In: *Methods in Molecular Biology* 840, pp. 177–188. ISSN: 10643745. DOI: 10.1007/978-1-61779-516-9_21. URL: https://link.springer.com/protocol/10.1007/978-1-61779-516-9{_}21.

Hutter, Carl R., Kerry A. Cobb, Daniel M. Portik, Scott L. Travers, Perry L. Wood, and Rafe M. Brown (2022). "FrogCap: A modular sequence capture probe-set for phylogenomics and population genetics for all frogs, assessed across multiple phylogenetic scales". In: *Molecular Ecology Resources* 22.3, pp. 1100–1119. ISSN: 1755-098X. DOI: 10.1111/1755-0998.13517. URL: https://onlinelibrary.wiley.com/doi/10.1111/1755-0998.13517.

Jennings, Mark R. and Galen B. Rathbun (2001). "Petrodromus tetradactylus". In: *Mammalian Species* 682.682, pp. 1–6. ISSN: 0076-3519. DOI: `10.1644/1545-1410(2001)682<0001:pt>2.0.co;2`.

Jong, Wilfried W. de, J. Timothy Gleaves, and Donald Boulter (1977). "Evolutionary changes of α-crystallin and the phylogeny of mammalian orders". In: *Journal of Molecular Evolution* 10.2, pp. 123–135. ISSN: 14321432. DOI: `10.1007/BF01751806`. URL: `https://link.springer.com/article/10.1007/BF01751806`.

Joordens, Josephine C.A., Craig S. Feibel, Hubert B. Vonhof, Anne S. Schulp, and Dick Kroon (2019). "Relevance of the eastern African coastal forest for early hominin biogeography". In: *Journal of Human Evolution* 131, pp. 176–202. ISSN: 00472484. DOI: `10.1016/j.jhevol.2019.03.012`. URL: `https://doi.org/10.1016/j.jhevol.2019.03.012`.

Katoh, K. and D. M. Standley (2013). "MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability". In: *Molecular Biology and Evolution* 30.4, pp. 772–780. ISSN: 0737-4038. DOI: `10.1093/molbev/mst010`. URL: `https://academic.oup.com/mbe/article-lookup/doi/10.1093/molbev/mst010`.

Kawasha, Boniface Kaumba (2003). "LUNDA GRAMMAR: A MORPHOSYNTACTIC AND SEMANTIC ANALYSIS". P.hD. Thesis. University of Oregon.

Kedves, M. (1971). "Presence de types sporomorphs importants dans les sediments prequaternaires Egyptiens". In: *Acta Botanica Academiae Scientarum Hungaricae* 17, pp. 371–378.

Kent, W. James, Robert Baertsch, Angie Hinrichs, Webb Miller, and David Haussler (2003). "Evolution's cauldron: Duplication, deletion and rearrangement in the mouse and human genomes". In: *Proceedings of the National Academy of Sciences of the United States of America* 100.20, pp. 11484–11489. ISSN: 00278424. DOI: `10.1073/pnas.1932072100`. URL: `www.soe.ucsc.edukent.`.

Kingdon, Jonathan (1997). *The Kingdon field guide to African mammals*. London : Academic Press.

— (2013). *"Chapter 4" in Mammals of Africa Volume I*. Ed. by Jonathan Kingdon. Bloomsbury Natural History, pp. 43–56. DOI: `10.5040/9781472926913`.

Kirilenko, Bogdan M, Chetan Munegowda, Ekaterina Osipova, David Jebb, Virag Sharma, Moritz Blumer, Ariadna E Morales, Alexis-Walid Ahmed, Dimitrios-Georgios Kontopoulos, Leon Hilgers, Kerstin Lindblad-Toh, Elinor K Karlsson, Zoonomia Consortium‡, Michael Hiller, Gregory Andrews, Joel C Armstrong, Matteo Bianchi, Bruce W Birren, Kevin R Bredemeyer, Ana M Breit, Matthew J Christmas, Hiram Clawson, Joana Damas, Federica Di Palma, Mark Diekhans, Michael X Dong, Eduardo Eizirik, Kaili Fan, Cornelia Fanter, Nicole M Foley,

Karin Forsberg-Nilsson, Carlos J Garcia, John Gatesy, Steven Gazal, Diane P Genereux, Linda Goodman, Jenna Grimshaw, Michaela K Halsey, Harris andrew J, Glenn Hickey, Michael Hiller, Allyson G Hindle, Robert M Hubley, Graham M Hughes, Jeremy Johnson, David Juan, Irene M Kaplow, Elinor K Karlsson, Kathleen C Keough, Bogdan Kirilenko, Klaus-Peter Koepfli, Jennifer M Korstian, Amanda Kowalczyk, Sergey V Kozyrev, Alyssa J Lawler, Colleen Lawless, Thomas Lehmann, Danielle L Levesque, Harris A Lewin, Xue Li, Abigail Lind, Kerstin Lindblad-Toh, Ava Mackay-Smith, Voichita D Marinescu, Tomas Marques-Bonet, Victor C Mason, Jennifer R S Meadows, Wynn K Meyer, Jill E Moore, Lucas R Moreira, Diana D Moreno-Santillan, Kathleen M Morrill, Gerard Muntané, William J. Murphy, Arcadi Navarro, Martin Nweeia, Sylvia Ortmann, Austin Osmanski, Benedict Paten, Nicole S Paulat, Pfenning andreas R, BaDoi N Phan, Katherine S Pollard, Henry E Pratt, David A Ray, Steven K Reilly, Jeb R Rosen, Irina Ruf, Louise Ryan, Oliver A Ryder, Pardis C Sabeti, Daniel E Schäffer, Aitor Serres, Beth Shapiro, Arian F A Smit, Mark S. Springer, Chaitanya Srinivasan, Cynthia Steiner, Jessica M Storer, Kevin A M Sullivan, Patrick F Sullivan, Elisabeth Sundström, Megan A Supple, Ross Swofford, Joy-El Talbot, Emma Teeling, Jason Turner-Maier, Alejandro Valenzuela, Franziska Wagner, Ola Wallerman, Chao Wang, Juehan Wang, Zhiping Weng, Aryn P Wilder, Morgan E Wirthlin, James R Xue, and Xiaomeng Zhang (2023). "Integrating gene annotation with orthology inference at scale." In: *Science (New York, N.Y.)* 380.6643, eabn3107. ISSN: 1095-9203. DOI: 10.1126/science.abn3107. URL: http://www.ncbi.nlm.nih.gov/pubmed/37104600.

Kocsis, LászÍ, Emmanuel Gheerbrant, Mustapha Mouflih, Henri Cappetta, Johan Yans, and Mbarek Amaghzaz (2014). "Comprehensive stable isotope investigation of marine biogenic apatite from the late Cretaceous-early Eocene phosphate series of Morocco". In: *Palaeogeography, Palaeoclimatology, Palaeoecology* 394, pp. 74–88. ISSN: 00310182. DOI: 10.1016/j.palaeo.2013.11.002.

Kozłowski, Jan, Marek Konarzewski, and Marcin Czarnoleski (2020). "Coevolution of body size and metabolic rate in vertebrates: a life-history perspective". In: *Biological Reviews* 95.5, pp. 1393–1417. ISSN: 1469185X. DOI: 10.1111/brv.12615.

Krásová, Jarmila, Ondřej Mikula, Radim Šumbera, Sylvie Horáková, Jan Robovský, Danila S. Kostin, Aleksey A. Martynov, Leonid A. Lavrenchenko, and Josef Bryja (2021). "The Rufous Sengi is not Elephantulus—Multilocus reconstruction of evolutionary history of sengis from the subfamily Macroscelidinae". In: *Journal of Zoological Systematics and Evolutionary Research* 59.4, pp. 918–932. ISSN: 14390469. DOI: 10.1111/jzs.12460.

Kumar, Sudhir and Sankar Subramanian (2002). "Mutation rates in mammalian genomes". In: *Proceedings of the National Academy of Sciences of the United States of America* 99.2, pp. 803–808. ISSN: 00278424. DOI: 10.1073/pnas.022629899. URL: www.pnas.orgcgidoi10.1073pnas.022629899.

Kunene, Caroline, Stefan H. Foord, Nikolaj Scharff, Thomas Pape, Jagoba Malumbres-Olarte, and Thinandavha C. Munyai (2022). "Ant Diversity Declines with Increasing Elevation along the Udzungwa Mountains, Tanzania". In: *Diversity* 14.4, pp. 1–11. ISSN: 14242818. DOI: 10.3390/d14040260.

Kyriazis, Christopher C., Robert K. Wayne, and Kirk E. Lohmueller (2021). "Strongly deleterious mutations are a primary determinant of extinction risk due to inbreeding depression". In: *Evolution Letters* 5.1, pp. 33–47. ISSN: 2056-3744. DOI: 10.1002/evl3.209. URL: https://academic.oup.com/evlett/article/5/1/33/6697634.

Landis, Michael J., Nicholas J. Matzke, Brian R. Moore, and John P. Huelsenbeck (2013). "Bayesian Analysis of Biogeography when the Number of Areas is Large". In: *Systematic Biology* 62.6, pp. 789–804. ISSN: 1076-836X. DOI: 10.1093/sysbio/syt040. URL: https://academic.oup.com/sysbio/article/62/6/789/1708738.

Lanfear, Robert, Paul B. Frandsen, April M. Wright, Tereza Senfeld, and Brett Calcott (2016). "PartitionFinder 2: New Methods for Selecting Partitioned Models of Evolution for Molecular and Morphological Phylogenetic Analyses". In: *Molecular Biology and Evolution* 34.3, msw260. ISSN: 0737-4038. DOI: 10.1093/molbev/msw260. URL: https://academic.oup.com/mbe/article-lookup/doi/10.1093/molbev/msw260.

Lawson, Lucinda P., José Alfredo Samaniego Castruita, James S. Haile, Cristiano Vernesi, Francesco Rovero, and Eline D. Lorenzen (2021). "Unraveling elephant-shrews: Phylogenetic relationships and unexpected introgression among giant sengis". In: *Molecular Phylogenetics and Evolution* 154, p. 107001. ISSN: 10959513. DOI: 10.1016/j.ympev.2020.107001. URL: https://doi.org/10.1016/j.ympev.2020.107001.

Lawson, Lucinda P., Cristiano Vernesi, Silvia Ricci, and Francesco Rovero (2013). "Evolutionary History of the Grey-Faced Sengi, Rhynchocyon udzungwensis, from Tanzania: A Molecular and Species Distribution Modelling Approach". In: *PLoS ONE* 8.8, pp. 1–10. ISSN: 19326203. DOI: 10.1371/journal.pone.0072506.

Leigh, Jessica W. and David Bryant (2015). "POPART: Full-feature software for haplotype network construction". In: *Methods in Ecology and Evolution* 6.9, pp. 1110–1116. ISSN: 2041210X. DOI: 10.1111/2041-210X.12410.

Li, Chenhong, Michael Hofreiter, Nicolas Straube, Shannon Corrigan, and Gavin J.P. Naylor (2013). "Capturing protein-coding genes across highly divergent species". In: *BioTechniques* 54.6, pp. 321–326. ISSN: 07366205. DOI: 10.2144/000114039.

Li, Heng and Richard Durbin (2009). "Fast and accurate short read alignment with Burrows-Wheeler transform". In: *Bioinformatics* 25.14, pp. 1754–1760. ISSN: 13674803. DOI: 10.1093/bioinformatics/btp324. URL: https://pubmed.ncbi.nlm.nih.gov/19451168/.

— (2010). "Fast and accurate long-read alignment with Burrows-Wheeler transform". In: *Bioinformatics* 26.5, pp. 589–595. ISSN: 13674803. DOI: 10.1093/bioinformatics/btp698. URL: https://pubmed.ncbi.nlm.nih.gov/20080505/.

— (2011). "Inference of human population history from individual whole-genome sequences". In: *Nature* 475.7357, pp. 493–496. ISSN: 00280836. DOI: 10.1038/nature10231. URL: www.nature.com/nature..

Li, Heng, Bob Handsaker, Alec Wysoker, Tim Fennell, Jue Ruan, Nils Homer, Gabor Marth, Goncalo Abecasis, and Richard Durbin (2009). "The Sequence Alignment/Map format and SAMtools". In: *Bioinformatics* 25.16, pp. 2078–2079. ISSN: 13674803. DOI: 10.1093/bioinformatics/btp352. URL: https://pubmed.ncbi.nlm.nih.gov/19505943/https://pubmed.ncbi.nlm.nih.gov/19505943/?dopt=Abstract.

Li, Heng and Nils Homer (2010). "A survey of sequence alignment algorithms for next-generation sequencing". In: *Briefings in Bioinformatics* 11.5, pp. 473–483. ISSN: 14675463. DOI: 10.1093/bib/bbq015.

Linder, H. Peter, Helen M. de Klerk, Julia Born, Neil D. Burgess, Jon Fjeldså, and Carsten Rahbek (2012). "The partitioning of Africa: statistically defined biogeographical regions in sub-Saharan Africa". In: *Journal of Biogeography* 39.7, pp. 1189–1205. ISSN: 03050270. DOI: 10.1111/j.1365-2699.2012.02728.x. URL: https://onlinelibrary.wiley.com/doi/10.1111/j.1365-2699.2012.02728.x.

Lukoschek, Vimoksalehi, J. Scott Keogh, and John C. Avise (2012). "Evaluating Fossil Calibrations for Dating Phylogenies in Light of Rates of Molecular Evolution: A Comparison of Three Approaches". In: *Systematic Biology* 61.1, p. 22. ISSN: 1076-836X. DOI: 10.1093/sysbio/syr075. URL: https://academic.oup.com/sysbio/article/61/1/22/1676748.

Marcovitz, Amir, Yatish Turakhia, Heidi I. Chen, Michael Gloudemans, Benjamin A. Braun, Haoqing Wang, and Gill Bejerano (2019). "A functional enrichment test for molecular convergent evolution finds a clear protein-coding signal in echolocating bats and whales". In: *Proceedings of the National Academy of Sciences of the United States of America* 116.42, pp. 21094–21103. ISSN: 10916490. DOI: 10.1073/pnas.1818532116. URL: www.pnas.org/cgi/doi/10.1073/pnas.1818532116.

Maricic, Tomislav, Mark Whitten, and Svante Pääbo (2010). "Multiplexed DNA sequence capture of mitochondrial genomes using PCR products". In: *PLoS ONE* 5.11, pp. 9–13. ISSN: 19326203. DOI: 10.1371/journal.pone.0014004.

Marivaux, Laurent, El Mabrouk Essid, Wissem Marzougui, Hayet Khayati Ammar, Sylvain Adnet, Bernard Marandat, Gilles Merzeraud, Rodolphe Tabuce, and Monique Vianey-Liaud (2014). "A new and primitive species of Protophiomys (Rodentia, Hystricognathi) from the late middle Eocene of Djebel el Kébar, Central Tunisia". In: *Palaeovertebrata* 38.1. ISSN: 00310247. DOI: 10.18563/pv.38.1.e2. URL: http://palaeovertebrata.com/articles/view/43.

Marshall, Charles R. (2008). "A simple method for bracketing absolute divergence times on molecular phylogenies using multiple fossil calibration points". In: *American Naturalist* 171.6, pp. 726–742. ISSN: 00030147. DOI: 10.1086/587523. URL: https://www.journals.uchicago.edu/doi/abs/10.1086/587523.

— (2019). *Using the Fossil Record to Evaluate Timetree Timescales*. DOI: 10.3389/fgene.2019.01049.

Martin, Marcel (2011). "Cutadapt removes adapter sequences from high-throughput sequencing reads". In: *EMBnet.journal* 17.1, p. 10. DOI: 10.14806/ej.17.1.200.

Mather, Niklas, Samuel M. Traves, and Simon Y.W. Ho (2020). "A practical introduction to sequentially Markovian coalescent methods for estimating demographic history from genomic data". In: *Ecology and Evolution* 10.1, pp. 579–589. ISSN: 20457758. DOI: 10.1002/ece3.5888.

Matzke, Nicholas J. (2014). "Model Selection in Historical Biogeography Reveals that Founder-Event Speciation Is a Crucial Process in Island Clades". In: *Systematic Biology* 63.6, pp. 951–970. ISSN: 1076-836X. DOI: 10.1093/sysbio/syu056. URL: https://academic.oup.com/sysbio/article/63/6/951/2847892.

McKusick, Victor A. (2007). *Mendelian Inheritance in Man and its online version, OMIM*. DOI: 10.1086/514346.

Meyer, Matthias and Martin Kircher (2010). "Illumina sequencing library preparation for highly multiplexed target capture and sequencing". In: *Cold Spring Harbor Protocols* 5.6. ISSN: 15596095. DOI: 10.1101/pdb.prot5448.

Minh, Bui Quang, Heiko A. Schmidt, Olga Chernomor, Dominik Schrempf, Michael D. Woodhams, Arndt Von Haeseler, Robert Lanfear, and Emma Teeling (2020). "IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the Genomic Era". In: *Molecular Biology and Evolution* 37.5, pp. 1530–1534. ISSN: 15371719. DOI: 10.1093/molbev/msaa015.

Mohandesan, Elmira, Camilla F. Speller, Joris Peters, Hans-Peter Uerpmann, Margarethe Uerpmann, Bea De Cupere, Michael Hofreiter, and Pamela A. Burger (2017). "Combined hybridization capture and shotgun sequencing for ancient DNA analysis of extinct wild and domestic dromedary camel". In: *Molecular Ecology Resources* 17.2, pp. 300–313. ISSN: 1755098X. DOI: 10.1111/1755-0998.12551. URL: https://onlinelibrary.wiley.com/doi/10.1111/1755-0998.12551.

Morley, Robert J. (2000). *Geological Evolution of Tropical Rainforests*. London: Wiley & Sons, p. 362.

Morley, Robert J. and Jonathan Kingdon (2013). "Africa's Environmental and Climatic Past". In: *Mammals of Africa Volume I*, pp. 43–56.

Murphy, William J., E. Eizirik, S. J. O'Brien, O. Madsen, M. Scally, C. J. Douady, E. Teeling, O. A. Ryder, Michael J. Stanhope, W. W. De Jong, and Mark S. Springer (2001a). "Resolution of the early placental mammal radiation using bayesian phylogenetics". In: *Science* 294.5550, pp. 2348–2351. ISSN: 00368075. DOI: 10.1126/science.1067179. URL: https://pubmed.ncbi.nlm.nih.gov/11743200/.

Murphy, William J., Eduardo Eizirik, Warren E. Johnson, Ya Ping Zhang, Oliver A. Ryder, and Stephen J. O'Brien (2001b). "Molecular phylogenetics and the origins of placental mammals". In: *Nature* 409.6820, pp. 614–618. ISSN: 00280836. DOI: 10.1038/35054550. URL: https://pubmed.ncbi.nlm.nih.gov/11214319/.

Nam, Kiwoong, Carina Mugal, Benoit Nabholz, Holger Schielzeth, Jochen B.W. Wolf, Niclas Backström, Axel Künstner, Christopher N. Balakrishnan, Heger andreas, Chris P. Ponting, David F. Clayton, and Hans Ellegren (2010). "Molecular evolution of genes in avian genomes". In: *Genome Biology* 11.6. ISSN: 14747596. DOI: 10.1186/gb-2010-11-6-r68.

Nikolaev, Sergey, Juan I. Montoya-Burgos, Elliott H. Margulies, NISC Comparative Sequencing Program, Jacques Rougemont, Bruno Nyffeler, and Stylianos E. Antonarakis (2007). "Early History of Mammals Is Elucidated with the ENCODE Multiple Species Sequencing Data". In: *PLoS Genetics* 3.1, e2. ISSN: 1553-7390. DOI: 10.1371/journal.pgen.0030002. URL: https://dx.plos.org/10.1371/journal.pgen.0030002.

Nishihara, Hidenori, Yoko Satta, Masato Nikaido, J. G.M. Thewissen, Michael J. Stanhope, and Norihiro Okada (2005). "A retroposon analysis of afrotherian phylogeny". In: *Molecular Biology and Evolution* 22.9, pp. 1823–1833. ISSN: 07374038. DOI: 10.1093/molbev/msi179.

Novacek, Michael (1984). "Evolutionary Stasis in the Elephant-Shrew, Rhynchocyon". In: pp. 4–22. DOI: 10.1007/978-1-4613-8271-3_2.

Novacek, Michael J. (1992). *Mammalian phytogeny: Shaking the tree*. DOI: 10.1038/356121a0. URL: https://www.nature.com/articles/356121a0.

Nursyifa, Casia, Anna Brüniche-Olsen, Genis Garcia-Erill, Rasmus Heller, and Albrechtsen anders (2022). "Joint identification of sex and sex-linked scaffolds in non-model organisms using low depth sequencing data". In: *Molecular Ecology Resources* 22.2, pp. 458–467. ISSN: 1755-098X. DOI: 10.1111/1755-0998.13491. URL: https://onlinelibrary.wiley.com/doi/10.1111/1755-0998.13491.

Olbricht, G and A Sliwa (2014). "Elephant shrews (sengis) - neither rodent nor shrew: A historical perspective". In: *Navorsing van die Nasional Museum Bloemfontein* 30.2, pp. 18–20.

Oldfield, Thomas (1897). "Exhibition of small mammals from North Nyasaland, with a note on the genus Petrodromus." In: *Proceedings of the Zoological Society of London*, pp. 430–436.

— (1910). "African mammals". In: *Annals and Magazine of Natural History 5*, pp. 83–92.

— (1918). "Notes on Petrodromus and Rhynchocyon". In: *The Annals and magazine of natural history 9*, pp. 364–370.

Olivier, Charlotte Anaïs, Adrian V. Jaeggi, Loren D. Hayes, and Carsten Schradin (2022). "Revisiting the components of Macroscelidea social systems: Evidence for variable social organization, including pair-living, but not for a monogamous mating system". In: *Ethology* 128.5, pp. 383–394. ISSN: 14390310. DOI: 10.1111/eth.13271.

Osipova, Ekaterina, Nikolai Hecker, and Michael Hiller (2019). "RepeatFiller newly identifies megabases of aligning repetitive sequences and improves annotations of conserved non-exonic elements". In: *GigaScience* 8.11, pp. 1–10. ISSN: 2047217X. DOI: 10.1093/gigascience/giz132.

Oxenham, Katherine H. and Michael R. Perrin (2009). "The Spatial Organization of the Four-Toed Elephant-Shrew (Petrodromus tetradactylus) in Tembe Elephant Park, KwaZulu-Natal, South Africa". In: *African Zoology* 44.2, pp. 171–180. ISSN: 1562-7020. DOI: 10.3377/004.044.0205.

Pääbo, Svante (1985). "Molecular cloning of Ancient Egyptian mummy DNA". In: *Nature* 314.6012, pp. 644–645. ISSN: 00280836. DOI: 10.1038/314644a0. URL: https://www.nature.com/articles/314644a0.

Paijmans, Johanna L.A., Sina Baleka, Kirstin Henneberger Last, and P C R Gold Buffer (2016). "In-solution Capture (Horn 2012 & Maricic 2010/revised2014)". In: 20.10 mL, pp. 1–5.

Paijmans, Johanna L.A., Axel Barlow, Kirstin Henneberger, Joerns Fickel, Michael Hofreiter, and Daniel W. G. Foerster (2020). "Ancestral mitogenome capture of the Southeast Asian banded linsang". In: *PLOS ONE* 15.6. Ed. by Tzen-Yuh Chiang, e0234385. ISSN: 1932-6203. DOI: 10.1371/journal.pone.0234385. URL: https://dx.plos.org/10.1371/journal.pone.0234385.

Paradis, Emmanuel (2018). "Analysis of haplotype networks: The randomized minimum spanning tree method". In: *Methods in Ecology and Evolution* 9.5, pp. 1308–1317. ISSN: 2041210X. DOI: 10.1111/2041-210X.12969.

Pekar, Stephen F and Robert M Deconto (2006). "High-resolution ice-volume estimates for the early Miocene: Evidence for a dynamic ice sheet in Antarctica". In:

Palaeogeography, Palaeoclimatology, Palaeoecology 231, pp. 101–109. DOI: `10.1016/j.palaeo.2005.07.027`. URL: `www.elsevier.com`.

Perrin, Mike and Galen B. Rathbun (2013). ""Order MACROSCELIDEA" in Mammals of Africa Volume I". In: ed. by Jonathan Kingdon. Bloomsbury Natural History, pp. 258–287. DOI: `10.5040/9781472926913`.

Pickford, Martin (2015). "Cenozoic geology of the northern Sperrgebiet, Namibia, accenting the Palaeogene". In: *Communications of the Geological Survey of Namibia* 16, pp. 10–104.

— (2020). "Two new rodents (Rodentia, Mammalia) from the late middle Eocene of Eocliff, Namibia". In: *Communications of the Geological Survey of Namibia* 22, pp. 21–46.

Pickford, Martin and Brigitte Senut (2008). *Geology and palaeobiology of the northern sperrgebiet: General conclusions and summary.* 20, pp. 555–573. ISBN: 9789994568765.

Plana, Vanessa (2004). "Mechanisms and tempo of evolution in the African Guineo-Congolian rainforest". In: *Philosophical Transactions of the Royal Society B: Biological Sciences*. Vol. 359. 1450. Royal Society, pp. 1585–1594. DOI: `10.1098/rstb.2004.1535`. URL: `https://royalsocietypublishing.org/`.

Potts, Richard (2013). *Hominin evolution in settings of strong environmental variability.* DOI: `10.1016/j.quascirev.2013.04.003`.

Rathbun, Galen B. (1979). "The social structure and ecology of elephant-shrews". In: *Fortschritte der Verhaltensforschung/Advances in Ethology* 20, pp. 1–75.

— (2009a). "AFROTHERIAN CONSERVATION". In: *Newsletter of the IUCN/SSC Afrotheria Special 1st Group Number* 7.7. ISSN: 02759306. DOI: `10.1109/PESC.1981.7083610`.

— (2009b). "Why is there discordant diversity in sengi (Mammalia: Afrotheria: Macroscelidea) taxonomy and ecology?" In: *African Journal of Ecology* 47.1, pp. 1–13. ISSN: 01416707. DOI: `10.1111/j.1365-2028.2009.01102.x`.

Rathbun, Galen B. and John P. Dumbacher (2015). "Home range and use of diurnal shelters by the Etendeka round-eared sengi, a newly discovered Namibian endemic desert mammal". In: *PeerJ* 2015.10. ISSN: 21678359. DOI: `10.7717/peerj.1302`.

Rathbun, Galen B. and C. D. Rathbun (2006). "Social structure of the bushveld sengi (Elephantulus intufi) in Namibia and the evolution of monogamy in the Macroscelidea". In: *Journal of Zoology* 269.3, pp. 391–399. ISSN: 0952-8369. DOI: `10.1111/j.1469-7998.2006.00087.x`. URL: `https://onlinelibrary.wiley.com/doi/10.1111/j.1469-7998.2006.00087.x`.

Rathbun, Galen B. and Peter F. Woodall (2002). "A bibliography of elephant-shrews or sengis (Macroscelidea)". In: *Mammal Review* 32.1, pp. 66–70. ISSN: 0305-1838.

DOI: `10.1046/j.1365-2907.2002.00092.x`. URL: `http://doi.wiley.com/10.1046/j.1365-2907.2002.00092.x`.

Raxworthy, Christopher J. and Brian Tilston Smith (2021). *Mining museums for historical DNA: advances and challenges in museomics*. DOI: `10.1016/j.tree.2021.07.009`. URL: `https://doi.org/10.1016/j.tree.2021.07.009`.

Ree, Richard H., Brian R. Moore, Campbell O. Webb, and Michael J. Donoghue (2005). "A LIKELIHOOD FRAMEWORK FOR INFERRING THE EVOLUTION OF GEOGRAPHIC RANGE ON PHYLOGENETIC TREES". In: *Evolution* 59.11, pp. 2299–2311. ISSN: 0014-3820. DOI: `10.1111/j.0014-3820.2005.tb00940.x`. URL: `https://onlinelibrary.wiley.com/doi/10.1111/j.0014-3820.2005.tb00940.x`.

Reis, M. d. dos and Z. Yang (2011). "Approximate Likelihood Calculation on a Phylogeny for Bayesian Estimation of Divergence Times". In: *Molecular Biology and Evolution* 28.7, pp. 2161–2172. ISSN: 0737-4038. DOI: `10.1093/molbev/msr045`. URL: `https://academic.oup.com/mbe/article-lookup/doi/10.1093/molbev/msr045`.

Rohland, Nadin and Michael Hofreiter (2007). "Ancient dna extraction from bones and teeth". In: *Nature Protocols* 2.7, pp. 1756–1762. ISSN: 17502799. DOI: `10.1038/nprot.2007.247`.

Rohland, Nadin, Heike Siedel, and Michael Hofreiter (2004). "Nondestructive DNA extraction method for mitochondrial DNA analyses of museum specimens". In: *BioTechniques* 36.5, pp. 814–821. ISSN: 07366205. DOI: `10.2144/04365st05`.

Ronquist, Fredrik (1997). "Dispersal-Vicariance Analysis: A New Approach to the Quantification of Historical Biogeography". In: *Systematic Biology* 46.1. Ed. by David Cannatella, pp. 195–203. ISSN: 1076-836X. DOI: `10.1093/sysbio/46.1.195`. URL: `http://academic.oup.com/sysbio/article/46/1/195/1685511`.

Ronquist, Fredrik, Maxim Teslenko, Paul van der Mark, Daniel L. Ayres, Aaron Darling, Sebastian Höhna, Bret Larget, Liang Liu, Marc A. Suchard, and John P. Huelsenbeck (2012). "MrBayes 3.2: Efficient Bayesian Phylogenetic Inference and Model Choice Across a Large Model Space". In: *Systematic Biology* 61.3, pp. 539–542. ISSN: 1076-836X. DOI: `10.1093/sysbio/sys029`. URL: `https://academic.oup.com/sysbio/article/61/3/539/1674894`.

Rovero, Francesco and Daniela W. De Luca (2007). "Checklist of mammals of the Udzungwa Mountains of Tanzania". In: *Mammalia* 71.1-2, pp. 47–55. ISSN: 00251461. DOI: `10.1515/MAMM.2007.015`.

Rovero, Francesco, Galen B. Rathbun, A. Perkin, T. Jones, D. O. Ribble, C. Leonard, R. R. Mwakisoma, and N. Doggart (2008). "A new species of giant sengi or elephant-shrew (genus Rhynchocyon) highlights the exceptional biodiversity

of the Udzungwa Mountains of Tanzania". In: *Journal of Zoology* 274.2, pp. 126–133. ISSN: 0952-8369. DOI: 10.1111/j.1469-7998.2007.00363.x. URL: https://onlinelibrary.wiley.com/doi/10.1111/j.1469-7998.2007.00363.x.

Salard-Cheboldaeff, M. (1990). "Intertropical African palynostratigraphy from Cretaceous to late quaternary times". In: *Journal of African Earth Sciences* 11.1-2, pp. 1–24. ISSN: 1464343X. DOI: 10.1016/0899-5362(90)90072-M.

Sallam, Hesham M. and Erik R. Seiffert (2016). "New phiomorph rodents from the latest Eocene of Egypt and the impact of Bayesian "clock"-based phylogenetic methods on estimates of basal hystricognath relationships and biochronology". In: *PeerJ* 2016.3, e1717. ISSN: 21678359. DOI: 10.7717/peerj.1717. URL: https://peerj.com/articles/1717.

Sallam, Hesham M and Erik R Seiffert (2020). "Revision of Oligocene 'Paraphiomys' and an origin for crown Thryonomyoidea (Rodentia: Hystricognathi: Phiomorpha) near the Oligocene–Miocene boundary in Africa". In: *Zoological Journal of the Linnean Society* 190.1, pp. 352–371. ISSN: 0024-4082. DOI: 10.1093/zoolinnean/zlz148. URL: https://academic.oup.com/zoolinnean/article/190/1/352/5691033.

Sánchez-Villagra, Marcelo R., Yuichi Narita, and Shigeru Kuratani (2007). *Thoracolumbar vertebral number: The first skeletal synapomorphy for afrotherian mammals*. DOI: 10.1017/S1477200006002258. URL: https://www.tandfonline.com/action/journalInformation?journalCode=tsab20.

Sansom, Robert S., Sarah E. Gabbott, and Mark A. Purnell (2010). "Non-random decay of chordate characters causes bias in fossil interpretation". In: *Nature* 463.7282, pp. 797–800. ISSN: 00280836. DOI: 10.1038/nature08745.

Sayyari, Erfan, James B. Whitfield, and Siavash Mirarab (2018). "DiscoVista: Interpretable visualizations of gene tree discordance". In: *Molecular Phylogenetics and Evolution* 122, pp. 110–115. ISSN: 10959513. DOI: 10.1016/j.ympev.2018.01.019. arXiv: 1709.09305.

Scally, Mark, Ole Madsen, Christophe J. Douady, Wilfried W. de Jong, Michael J. Stanhope, and Mark S. Springer (2001). "Molecular Evidence for the Major Clades of Placental Mammals". In: *Journal of Mammalian Evolution* 8.4, pp. 239–277. ISSN: 10647554. DOI: 10.1023/A:1014446915393. URL: http://link.springer.com/10.1023/A:1014446915393.

Schriml, Lynn M., Elvira Mitraka, James Munro, Becky Tauber, Mike Schor, Lance Nickle, Victor Felix, Linda Jeng, Cynthia Bearer, Richard Lichenstein, Katharine Bisordi, Nicole Campion, Brooke Hyman, David Kurland, Connor Patrick Oates, Siobhan Kibbey, Poorna Sreekumar, Chris Le, Michelle Giglio, and Carol Greene

(2019). "Human Disease Ontology 2018 update: Classification, content and workflow expansion". In: *Nucleic Acids Research* 47.D1, pp. D955–D962. ISSN: 13624962. DOI: 10.1093/nar/gky1032. URL: http://ceur-ws.org/Vol-1515/demo6.pdf.

Seabold, S and J Perktold (2010). *No statsmodels: Econometric and statistical modeling with python. In 9th Python in Science Conference.*

Seiffert, Erik R. (2007). "A new estimate of afrotherian phylogeny based on simultaneous analysis of genomic, morphological and fossil evidence". In: *BMC Evolutionary Biology* 7.1. ISSN: 14712148. DOI: 10.1186/1471-2148-7-224.

— (2010). "The oldest and youngest records of afrosoricid placentals from the Fayum depression of Northern Egypt". In: *Acta Palaeontologica Polonica* 55.4, pp. 599–616. ISSN: 05677920. DOI: 10.4202/app.2010.0023.

Senut, Brigitte and Martin Pickford (2021). "Micro-cursorial mammals from the late Eocene tufas at Eocliff, Namibia". In: *Communications of the Geological Survey of Namibia* 23, pp. 90–160.

Shoshani, Jeheskel (1986). "Mammalian phylogeny: comparison of morphological and molecular results." In: *Molecular Biology and Evolution* 3.3, pp. 222–242. ISSN: 1537-1719. DOI: 10.1093/oxfordjournals.molbev.a040389. URL: https://academic.oup.com/mbe/article/3/3/222/982658/Mammalian-phylogeny-comparison-of-morphological.

Shoshani, Jeheskel and Malcolm C. McKenna (1998). "Higher Taxonomic Relationships among Extant Mammals Based on Morphology, with Selected Comparisons of Results from Molecular Data". In: *Molecular Phylogenetics and Evolution* 9.3, pp. 572–584. ISSN: 10557903. DOI: 10.1006/mpev.1998.0520.

Smit, Hanneline A. (2008). "A NEW SPECIES OF ELEPHANT-SHREW (AFROTHERIA : MACROSCELIDEA : ELEPHANTULUS) FROM SOUTH AFRICA". In: 89.5, pp. 1257–1269.

Smit, Hanneline A., B. Jansen van Vuuren, P. C.M. O'Brien, M. Ferguson-Smith, F. Yang, and T. J. Robinson (2011). "Phylogenetic relationships of elephant-shrews (Afrotheria, Macroscelididae)". In: *Journal of Zoology* 284.2, pp. 133–143. ISSN: 09528369. DOI: 10.1111/j.1469-7998.2011.00790.x.

Soria, Carmen D., Michela Pacifici, Moreno Di Marco, Sarah M. Stephen, and Carlo Rondinini (2021). "COMBINE: a coalesced mammal database of intrinsic and extrinsic traits". In: *Ecology* 102.6, e03344. ISSN: 19399170. DOI: 10.1002/ecy.3344. URL: https://onlinelibrary.wiley.com/doi/10.1002/ecy.3344.

Springer, Mark S. (2022). "Afrotheria". In: *Current Biology* 32.5, R205–R210. ISSN: 18790445. DOI: 10.1016/j.cub.2022.02.001.

Springer, Mark S., Gregory C. Cleven, Ole Madsen, Wilfried W. De Jong, Victor G. Waddell, Heather M. Amrine, and Michael J. Stanhope (1997). "Endemic African

mammals shake the phylogenetic tree". In: *Nature* 388.6637, pp. 61–64. ISSN: 00280836. DOI: 10.1038/40386.

Springer, Mark S. and William J. Murphy (2007). *Mammalian evolution and biomedicine: New views from phylogeny*. DOI: 10.1111/j.1469-185X.2007.00016.x. URL: https://pubmed.ncbi.nlm.nih.gov/17624960/.

Springer, Mark S., Anthony V. Signore, Johanna L.A. Paijmans, Jorge Vélez-Juarbe, Daryl P. Domning, Cameron E. Bauer, Kai He, Lorelei Crerar, Paula F. Campos, William J. Murphy, Robert W. Meredith, John Gatesy, Eske Willerslev, Ross D.E. MacPhee, Michael Hofreiter, and Kevin L. Campbell (2015). "Interordinal gene capture, the phylogenetic position of Steller's sea cow based on molecular and morphological data and the macroevolutionary history of Sirenia". In: *Molecular Phylogenetics and Evolution* 91, pp. 178–193. ISSN: 10959513. DOI: 10.1016/j.ympev.2015.05.022. URL: http://dx.doi.org/10.1016/j.ympev.2015.05.022.

Springer, Mark S., Michael J. Stanhope, Ole Madsen, and Wilfried W. De Jong (2004). *Molecules consolidate the placental mammal tree*. DOI: 10.1016/j.tree.2004.05.006. URL: https://pubmed.ncbi.nlm.nih.gov/16701301/.

Stamatakis, Alexandros (2014). "RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies". In: *Bioinformatics* 30.9, pp. 1312–1313. ISSN: 1460-2059. DOI: 10.1093/bioinformatics/btu033. URL: https://academic.oup.com/bioinformatics/article-lookup/doi/10.1093/bioinformatics/btu033.

Stanhope, Michael J., Victor G. Waddell, Ole Madsen, Wilfried De Jong, S. Blair Hedges, Gregory C. Cleven, Diana Kao, and Mark S. Springer (1998). "Molecular evidence for multiple origins of Insectivora and for a new order of endemic African insectivore mammals". In: *Proceedings of the National Academy of Sciences of the United States of America* 95.17, pp. 9967–9972. ISSN: 00278424. DOI: 10.1073/pnas.95.17.9967. URL: www.pnas.org..

Steffen, Raphael, Lynn Ogoniak, Norbert Grundmann, Anna Pawluchin, Oliver Soehnlein, and Jürgen Schmitz (2022). "paPAML: An Improved Computational Tool to Explore Selection Pressure on Protein-Coding Sequences". In: *Genes* 13.6. ISSN: 20734425. DOI: 10.3390/genes13061090. URL: https://pubmed.ncbi.nlm.nih.gov/35741852/https://pubmed.ncbi.nlm.nih.gov/35741852/?utm{_}source=gquery.

Stelzer, Gil, Naomi Rosen, Inbar Plaschkes, Shahar Zimmerman, Michal Twik, Simon Fishilevich, Tsippi Iny Stein, Ron Nudel, Iris Lieder, Yaron Mazor, Sergey Kaplan, Dvir Dahary, David Warshawsky, Yaron Guan-Golan, Asher Kohn, Noa Rappaport, Marilyn Safran, and Doron Lancet (2016). "The GeneCards Suite: From Gene Data Mining to Disease Genome Sequence Analyses". In: *Current Protocols in*

Bioinformatics 54.1, pp. 1.30.1–1.30.33. ISSN: 1934-3396. DOI: 10.1002/cpbi.5. URL: https://currentprotocols.onlinelibrary.wiley.com/doi/10.1002/cpbi.5.

Stevens, Nancy J., Patrick M. O'Connor, Cassy Mtelela, and Eric M. Roberts (2022). "Macroscelideans (Myohyracinae and Rhynchocyoninae) from the late Oligocene Nsungwe formation of the Rukwa Rift Basin, southwestern Tanzania". In: *Historical Biology* 34.4, pp. 604–610. ISSN: 10292381. DOI: 10.1080/08912963.2021.1938565. URL: https://doi.org/10.1080/08912963.2021.1938565.

Straube, Nicolas, Mariana L. Lyra, Johanna L.A. Paijmans, Michaela Preick, Nikolas Basler, Johannes Penner, Mark-Oliver Rödel, Michael V. Westbury, Célio F. B. Haddad, Axel Barlow, and Michael Hofreiter (2021). "Successful application of ancient DNA extraction and library construction protocols to museum wet collection specimens". In: *Molecular Ecology Resources* 21.7, pp. 2299–2315. ISSN: 1755-098X. DOI: 10.1111/1755-0998.13433. URL: https://onlinelibrary.wiley.com/doi/10.1111/1755-0998.13433.

Suarez, Hernando G., Bjoern E. Langer, Pradnya Ladde, and Michael Hiller (2017). "ChainCleaner improves genome alignment specificity and sensitivity". In: *Bioinformatics* 33.11, pp. 1596–1603. ISSN: 14602059. DOI: 10.1093/bioinformatics/btx024.

Sun, Yan Bo, Zi Jun Xiong, Xue Yan Xiang, Shi Ping Liu, Wei Wei Zhou, Xiao Long Tu, Li Zhong, Lu Wang, Dong Dong Wu, Bao Lin Zhang, Chun Ling Zhu, Min Min Yang, Hong Man Chen, Fang Li, Long Zhou, Shao Hong Feng, Chao Huang, Guo Jie Zhang, David Irwin, David M. Hillis, Robert W. Murphy, Huan Ming Yang, Jing Che, Jun Wang, and Ya Ping Zhang (2015). "Whole-genome sequence of the Tibetan frog Nanorana parkeri and the comparative evolution of tetrapod genomes". In: *Proceedings of the National Academy of Sciences of the United States of America* 112.11, E1257–E1262. ISSN: 10916490. DOI: 10.1073/pnas.1501764112.

Tabuce, Rodolphe, Robert J. Asher, and Thomas Lehmann (2008). "Afrotherian mammals: A review of current data". In: *Mammalia* 72.1, pp. 2–14. ISSN: 00251461. DOI: 10.1515/MAMM.2008.004.

Tabuce, Rodolphe, Laurent Marivaux, Mohammed Adaci, Mustapha Bensalah, Jean Louis Hartenberger, Mohammed Mahboubi, Fateh Mebrouk, Paul Tafforeau, and Jean Jacques Jaeger (2007). "Early Tertiary mammals from North Africa reinforce the molecular Afrotheria clade". In: *Proceedings of the Royal Society B: Biological Sciences* 274.1614, pp. 1159–1166. ISSN: 14712970. DOI: 10.1098/rspb.2006.0229. URL: https://pubmed.ncbi.nlm.nih.gov/17329227/.

Townsend, Jeffrey P. and Christoph Leuenberger (2011). "Taxon sampling and the optimal rates of evolution for phylogenetic inference". In: *Systematic Biology* 60.3, pp. 358–365. ISSN: 10635157. DOI: 10.1093/sysbio/syq097.

Trauth, Martin H., Mark A. Maslin, Alan Deino, and Manfred R. Strecker (2005). "Climate change: Late cenozoic moisture history of east Africa". In: *Science* 309.5743, pp. 2051–2053. ISSN: 00368075. DOI: 10.1126/science.1112964.

Valk, Tom van der, David Díez-del Molino, Tomas Marques-Bonet, Katerina Guschanski, and Love Dalén (2019). "Historical Genomes Reveal the Genomic Consequences of Recent Population Decline in Eastern Gorillas". In: *Current Biology* 29.1, 165–170.e6. ISSN: 09609822. DOI: 10.1016/j.cub.2018.11.055.

Van Couvering, John A. and Eric Delson (2020). "African Land Mammal Ages". In: *Journal of Vertebrate Paleontology* 40.5, e1803340. ISSN: 0272-4634. DOI: 10.1080/02724634.2020.1803340. URL: https://www.tandfonline.com/doi/full/10.1080/02724634.2020.1803340.

Visendi, Paul (2022). "De Novo Assembly of Linked Reads Using Supernova 2.0". In: *Methods in Molecular Biology*. Vol. 2443. Humana Press Inc., pp. 233–243. DOI: 10.1007/978-1-0716-2067-0_12. URL: https://link.springer.com/protocol/10.1007/978-1-0716-2067-0{_}12.

White, C. N. M and W.F.H. Ansell (1966). "A list of Livale and Lunda mammal names". In: *The Puku* 4, pp. 181–185.

Wildman, Derek E., Monica Uddin, Juan C. Opazo, Guozhen Liu, Vincent Lefort, Stephane Guindon, Olivier Gascuel, Lawrence I. Grossman, Roberto Romero, and Morris Goodman (2007). "Genomics, biogeography and the diversification of placental mammals". In: *Proceedings of the National Academy of Sciences of the United States of America* 104.36, pp. 14395–14400. ISSN: 00278424. DOI: 10.1073/pnas.0704342104. URL: /pmc/articles/PMC1958817//pmc/articles/PMC1958817/?report=abstracthttps://www.ncbi.nlm.nih.gov/pmc/articles/PMC1958817/.

Wimsatt, W. A. (1975). "Some comparative aspects of implantation". In: *Biology of Reproduction* 12.1, pp. 1–40. ISSN: 00063363. DOI: 10.1095/biolreprod12.1.1.

Woodall, Peter F. and R. I. Mackie (1987). "Caecal size and function in the rock elephant shrew Elephantulus myurus (insectivora, macroscelididae) and the namaqua rock mouse Aethomys namaquensis (Rodentia, Muridae)". In: *Comparative Biochemistry and Physiology – Part A: Physiology* 87.2, pp. 311–314. ISSN: 03009629. DOI: 10.1016/0300-9629(87)90127-7.

Yang, Z. (2007). "PAML 4: Phylogenetic Analysis by Maximum Likelihood". In: *Molecular Biology and Evolution* 24.8, pp. 1586–1591. ISSN: 0737-4038. DOI: 10.1093/molbev/msm088. URL: https://academic.oup.com/mbe/article-lookup/doi/10.1093/molbev/msm088.

Zhang, Chao, Maryam Rabiee, Erfan Sayyari, and Siavash Mirarab (2018). "ASTRAL-III: Polynomial time species tree reconstruction from partially resolved gene

trees". In: *BMC Bioinformatics* 19.Suppl 6, pp. 15–30. ISSN: 14712105. DOI: 10.1186/s12859-018-2129-y. URL: http://dx.doi.org/10.1186/s12859-018-2129-y.

Zimmermann, Juergen, Mehrdad Hajibabaei, David C. Blackburn, James Hanken, Elizabeth Cantin, Janos Posfai, and Thomas C. Evans (2008). "DNA damage in preserved specimens and tissue samples: A molecular assessment". In: *Frontiers in Zoology* 5.18, p. 18. ISSN: 17429994. DOI: 10.1186/1742-9994-5-18. URL: http://www.frontiersinzoology.com/content/5/1/18http://www.frontiersinzoology.com/content/5/1/18.