School of Doctoral Studies in Biological Sciences
University of South Bohemia in České Budějovice
Faculty of Science

Diversity, evolution, and distribution of selected African rodents, with focus on East African savannahs

Ph.D. Thesis

Mgr. Vladimír Mazoch

Supervisor: doc. Mgr. Radim Šumbera, Ph.D.
Institute of Zoology, University of South Bohemia,
České Budějovice, the Czech Republic

Co-supervisor: doc. Mgr. et Mgr. Josef Bryja Ph.D.
Institute of Vertebrate Biology, the Czech Academy of Sciences,
Brno, the Czech Republic

České Budějovice 2018

This thesis should be cited as:

Mazoch, V., 2018: Diversity, evolution, and distribution of selected African rodents, with focus on East African savannahs. Ph.D. Thesis Series, No. 13. University of South Bohemia, Faculty of Science, School of Doctoral Studies in Biological Sciences, České Budějovice, Czech Republic, 117 pp.

Annotation

The thesis describes genetic diversity and phylogeography in several species of African rodents. It examines cryptic diversity, origin of these taxa together with colonization routes, refugia, barriers to gene flow and other historical factors influencing their current geographical distribution.

Molecular genetic analyses detected considerably higher diversity of small mammals than previously expected. Several candidates for new species have been suggested and many new localities for known species have been discovered. Observed phylogeographic structure revealed geographical barriers to gene flow and current contact zones between neighboring taxonomical units (Zambezi-Kafue river system, Rukwa Rift, Eastern Arc Mountains, Limpopo river valley, etc.). Molecular dating, evidence for expansion and contraction of population size, and species distribution modeling suggest connection of main lineage splits with rapid fluctuations in forest/savanna biomes during the Late Pleistocene.

Declaration [in Czech]

Prohlašuji, že svoji disertační práci jsem vypracoval samostatně pouze s použitím pramenů a literatury uvedených v seznamu citované literatury.

Prohlašuji, že v souladu s § 47b zákona č. 111/1998 Sb. v platném znění souhlasím se zveřejněním své disertační práce, a to v úpravě vzniklé vypuštěním vyznačených částí archivovaných Přírodovědeckou fakultou elektronickou cestou ve veřejně přístupné části database STAG provozované Jihočeskou univerzitou v Českých Budějovicích na jejích internetových stránkách, a to se zachováním mého autorského práva k odevzdanému textu této kvalifikační práce.

Souhlasím dále s tím, aby toutéž elektronickou cestou byly v souladu s uvedeným ustanovením zákona č. 111/1998 Sb. zveřejněny posudky školitele a oponentů práce i záznam o průběhu a výsledku obhajoby kvalifikační práce.

Rovněž souhlasím s porovnáním textu mé kvalifikační práce s databází kvalifikačních prací Theses.cz provozovanou Národním registrem vysokoškolských kvalifikačních prací a systémem na odhalování plagiátů, provozovanou Národním registrem vysokoškolských kvalifikačních prací a systémem na odhalování plagiátů.

České Budějovice, ...

.....

Mgr. Vladimír Mazoch

This thesis originated from a partnership of Faculty of Science, University of South Bohemia; and Institute of Vertebrate Biology, Czech Academy of Sciences.

Financial support

This thesis was financially supported by the projects of the Czech Science Foundation, nos. P506/10/0983 and 15-20229S.

Acknowledgements

The completion of this thesis could not have been possible without the support, assistance, and patience of many people who have accompanied me in recent years as I have worked on this thesis. To all of them, whose names are to be listed here, I wish to express my sincere thanks.

I would like to express my gratitude to my supervisor Radim Šumbera for professional guidance, encouragement, and advice he has provided throughout my time as his student, to Pepa Bryja and Ondra Mikula for their constructive suggestions and insightful comments during the writing of the manuscript. I wish to thank to my friend and colleague Hana Konvičková for introduction to lab work and phylogenetic analyses. I am thankful to Molly McDonough and Sophie Gryseels for providing me an opportunity to work with them on their respective manuscripts.

I would like to thank my fellow doctoral students Mates Lövy, Honza Šklíba, Ema Hrouzková, Honza Zima, Ondra Kott, Džája Smiešková, Táňa Aghová, and once again Sophie for their feedback, cooperation, and of course friendship, especially during field expeditions.

Field work would not be possible without our local colleagues, namely Yonas Meheretu, Kiros Welegerima, Abdul Katakweba, and Ramos Tseu. All of them and many others were not only great field workers but provided many valuable advices during writing of the manuscripts.

A very special gratitude goes out to teachers and staff at the University of South Bohemia and Research Facility of Institute of Vertebrate Biology in Studenec for providing fascinating lectures and access to the laboratory and research facilities.

I am also grateful to my significant other, my family, and friends for their support and encouragement throughout my prolonged studies.

List of papers and author's contributions

The thesis is based on the following papers:

I. **Mazoch V.**, Mikula O., Bryja J., Konvičková H., Russo I. R., Verheyen E. & Šumbera R. (2018). Phylogeography of a widespread sub-Saharan murid rodent *Aethomys chrysophilus*: the role of geographic barriers and paleoclimate in the Zambezian bioregion. *Mammalia*, 82(4), 373-387. (IF 2017 = 0.714).

Vladimír Mazoch (VM) provided, together with Josef Bryja and Radim Šumbera, substantial contribution to conception and design of the manuscript, drafted first version of the manuscript to the coauthors, revised critically important content, performed laboratory work, collaborate on statistical analysis, and contributed to field work.

II. McDonough M. M., Šumbera R., **Mazoch V.**, Ferguson A. W., Phillips C. D., & Bryja J. (2015). Multilocus phylogeography of a widespread savanna-woodland-adapted rodent reveals the influence of Pleistocene geomorphology and climate change in Africa's Zambezi region. *Molecular Ecology,* 24(20), 5248-5266. (IF 2017 = 6.131).

VM was involved in data collection, ecological niche modelling analysis and interpretation, and provided a revision of the manuscript.

III. Bryja J., Mikula O., Šumbera R., Meheretu Y., Aghová T., Lavrenchenko L. A., **Mazoch V.**, Oguge N., Mbau J. S., Welegerima K., Amundala N., Colyn M., Leirs H. & Verheyen E. (2014). Pan-African phylogeny of *Mus* (subgenus *Nannomys*) reveals one of the most successful mammal radiations in Africa. *BMC Evolutionary Biology*, 14(1), 256. (IF 2017 = 3.027).

VM contributed to data analyses, fieldwork, and contributed to writing of the manuscript.

IV. Bryja J., **Mazoch V.**, Patzenhauerová H., Mateke C., Zima J. Jr, Šklíba J. & Šumbera R. (2012). Revised occurrence of rodents from the tribe Praomyini (Muridae) in Zambia based on mitochondrial DNA analyses: implications for biogeography and conservation. *Folia Zoologica*, 61(3/4), 268-283. (IF 2017 = 0.467).

VM performed laboratory work, contributed to preparation, analysis of the dataset, manuscript editing, and manuscript review.

Content

General Introduction

1. Introduction

1.1. General background

An ongoing boom in studies combining molecular genetics and geographical data provides a capable tool for the understanding phylogenetic diversity and historical processes responsible for the current distribution of taxonomical units. These topics are of paramount importance especially in neglected areas across the globe. For example, in sub-Saharan Africa, descriptions of genetic diversity in, e.g. ungulates (reviewed in Lorenzen et al. 2012, and supplemented by Fennessy et al. 2016, Pedersen et al. 2018, Smitz et al. 2013), carnivores (Barnett et al. 2014, Bertola et al. 2016, Dubach et al. 2005), and monkeys (Haus et al. 2013, Zinner et al. 2013), helped to identify historical refugia and migration barriers in larger mammals; factors with direct implication for biodiversity conservation.

This boom is further fueled by a substantial improvement in the accuracy of predictions of potential geographic distribution through habitat modelling (Phillips et al. 2006). Climatic models used on the set of selected bioclimatic variables can predict not only the current potential distribution (Fick and Hijmans 2017), but through paleoclimatic modelling (Otto-Bliesner et al. 2006) geographic range shifts of a given species can be predicted throughout time as a consequence of the climatic changes.

Despite a remarkable advancement in the description of rodent diversity using phylogenetic methods in recent years (Fabre et al. 2012, Steppan and Schenk 2017, Monadjem et al. 2015), we are still far from fully understanding how the historical processes affected the current distribution of biota on the African continent. A stunning number of 158 completely new and almost 200 cryptic mammalian

species were discovered in Afrotropics from 2005 to 2018 (Burgin et al. 2018) and these numbers are probably still highly underestimated (see Bryja, Mikula et al. 2014b, Demos et al. 2014).

Nature of these species needs to be first properly defined and understood, as biodiversity certainly depends on the adopted species concept. Traditional typological or biological concepts are not always applicable for species delimitation in small mammals due to numerous cases of convergent evolution, mtDNA introgression (just to point out one recent example of adaptive introgression of mtDNA in rodents from Ethiopian highlands: Bryja et al. 2018), high diversity of cryptic species (Demos et al. 2014) and the difficulties in verification of reproductive isolation in allopatric species (Russo et al. 2006). Coupled with the rapidly increasing amount of genetic data from free-living populations, these concepts need to be complemented by species concepts based either on the genetic (Baker and Bradley 2006) or phylogenetic (Groves 2013) approach, creating thus the so-called integrative taxonomic approach. Genetic approaches often detect cryptic diversity within evolutionary lineages and provides important data influencing future conservation decisions (Groves et al. 2017).

1.2. East(ern) Africa

Region of East or eastern Africa forms a wide belt along the coast of the Indian Ocean. Definition of this specific subregion of sub-Saharan Africa by geography or geopolitics varies considerably between sources, but demarcation as referred in the United Nations Statistics Division scheme of geographic regions (https://unstats.un.org/unsd/methodology/m49/) is matching our sampling effort very closely. So, when text refers to East or eastern Africa, it means all countries highlighted in Figure 1, excluding islands.

During 15 years of our studies, we were able to cover significant part of East Africa by either direct fieldwork in several East African countries (Malawi, Zambia, Kenya, Tanzania, Mozambique, and Ethiopia) or by obtaining tissue samples from collaborators from various institutions across the world. Our dataset allowed us to conduct studies covering a significant part of the distributional range of selected species.

Factors influencing biodiversity in the region always work as an interconnected network, and can explain the observed geographical distribution of taxonomical units only when considered together and in a broader context. Formation of the East African Rift System (EARS) have created mountain ranges as barriers to gene flow. Uplifting of these mountain ranges changed water availability in the region, altered courses of many rivers, and led to formation of lakes. Another example of complex factors are rivers and their adjacent gorges. River course itself works as a direct barrier to gene flow, but in combination with inevitable habitat change along the riverbed, the barrier can be widened significantly by unsuitable environment. Together with climatic fluctuations caused by planetary mechanics, East Africa creates tumultuous environment for any species and as well for researchers attempting to untangle riddles in the specific scientific fields.

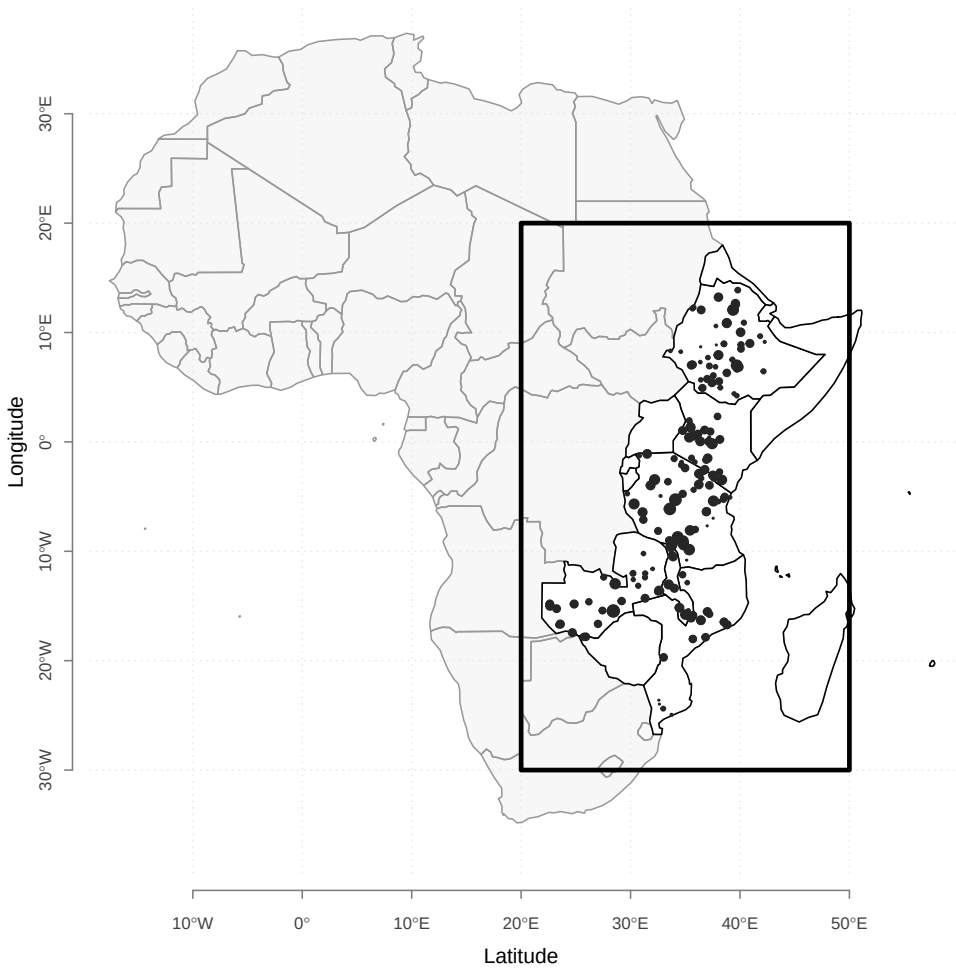


Figure 1: Map of East Africa (countries with black outline and white fill) as defined by the United Nations Statistics Division scheme of geographic regions. Trapping sites for all our field expeditions is shown with dark grey dots. Size of the dots reflect the number of trapped specimens per site. Area inside black rectangle will be shown in subsequent maps. All maps presented in the thesis were created using R statistical software (R Core Team 2018) and R Markdown ⟨http://rmarkdown.rstudio.com⟩.

1.3. The East African Rift System

The East African Rift System (EARS), the Earth's largest intra-continental rift system, extends from the Red Sea to Mozambique. The EARS shows a general North–South trend and shows up at the surface as a series of several thousand kilometers long aligned successions of adjacent individual tectonic basins (rift valleys), separated from each other by uplifted shoulders (Chorowicz 2005). The rift valleys form two main lines, the eastern and western branches of the EARS. Multiple Rifts are recognized within branches, with Albertine Rift and Gregory Rift being the most prominent (Figure 2).

The EARS transects the high-elevation (>1,000m) Ethiopian and East African plateaus that together form part of the 6,000-km-long topographic anomaly referred to as the African superswell (Nyblade and Robinson 1994). The uplift of eastern African topography and the onset of Antarctic glaciation led to decrease in rainfall, which led to strong aridification of East African climate (Sepulchre et al. 2006).

The uplift is structurally and magmatically controlled, creating complex relief and drainage conditions that are highly variable through time. The tectonic activity forming the current shape of East Africa started about 45 Mya with arising of the first volcanoes in Ethiopian Rift. The uplift itself started in the early Miocene and major uplifting around 15 Mya established Ethiopian plateau (Wolfenden et al. 2004). Multiple ages of rifting in East Africa are recognized but the majority of fault activity, structural growth, subsidence, and associated uplift of East Africa seem to have occurred in the last 5–9 Mya, particularly in the last 1–2 Mya, and is enduring to present (Roberts et al. 2012). However, Rift valley formation occurred in a number of stages on the North-South axis. Rifting in Kenya started as early as 23 Mya, northern part of Western branch followed (12 Mya) and its Southern counterpart was

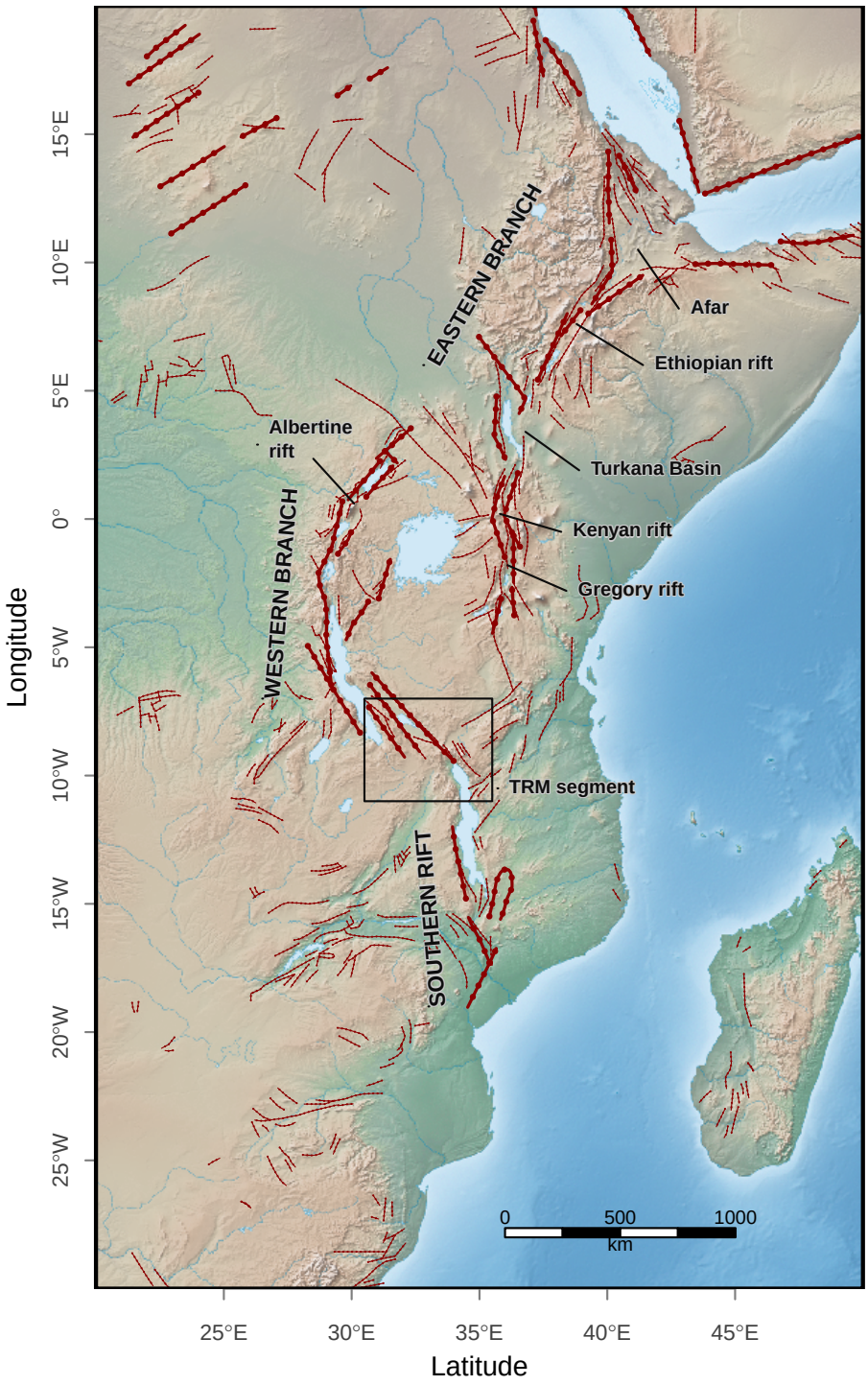


Figure 2: Hypsographic map of the East African Rift System. The main faults are marked in thick red lines. Thinner red dot-ted lines show other tectonic contacts. TRM stands for Tan-ganyika Rukwa Malawi segment.

next (7 Mya). The processes continued to the Plio-Pleistocene with and Rukwa (6 Mya) and Usangu (3 Mya) rifts (Macgregor 2015).

Physiogeographic processes associated with rifting as tectonic and volcanic activity, reduction of vegetation (intensified in the Late Pliocene/Early Pleistocene), erosion, and fluvial dynamics contributed to periodical fragmentation of habitats and led to various speciation events (O'Brien and Peters 1999).

1.4. Climatic fluctuations

Long-term climate change is driven primarily by tectonic activity. The global decreases of temperature in East Africa, increase of aridity, and consequent shift from wet to more arid environments started around mid-Miocene (16-13 Mya) with the uplift of the EARS (Zachos et al. 2001). This wide range climatic event promoted extensive spread of open biomes and eventually led to the final appearance of the Sahara Desert between 7 Mya and 2.5 Mya (Feakins et al. 2013).

It seems the heterogenous geology and climate of East Africa created periods of highly variable local climate. During Neogene and Quaternary, i.e. period when most current mammalian species evolved, the global shift to aridity was irregularly interrupted by short periods of rapid alteration between wet and arid conditions known as extreme climate variability (Maslin et al. 2014). The most prominent periods of the climate variability correlate with emergence of deep lakes (Trauth et al. 2005, 2007) and also with major global climatic episodes: the onset of Northern Hemisphere Glaciation (2.7-2.5 Mya), intensification of the Walker Circulation (1.9-1.7 Mya), and the Mid-Pleistocene (1.0-0.7 Mya) Revolution (Gornitz 2009).

Later in Quaternary, over the past 400 000 years, orbitally driven glacial variability exhibits pattern with periods near 23 000, 41 000,

and peaking in 100 000 years (Imbrie et al. 1992). Synchronization of these cycles led to intensification of the climate change. In the northern hemisphere, expansion of ice sheets led to cold dry climate and retreat of wooded vegetation in favor of grasslands. Opposite conditions during wet and humid inter-glacials favored expansion of forests and shrinking of grasslands. Widely used climatic models allow predicting fluctuations of habitats during the Last Glacial Maximum (LGM; ~22 000 years ago) and the Last Interglacial (LIG; 130 000 to 116 000 years ago) periods (Fick and Hijmans 2017, Otto-Bliesner et al. 2006). Obtained results need to be interpreted with caution, as specifics of East African historical conditions should be taken into account.

Africa presumably experienced a similar set of oscillations between warm humid and cool dry climate, followed by expansions or contractions of tropical forests and savannahs (de Menocal 1995, 2004), but glaciation of the southern hemisphere was less extensive and in East Africa rather different scenario occurred. Lake Malawi region (and on similar scale Lake Tanganyika) experienced condition of extreme aridity episodically between 135 and 70 ka (Scholz et al. 2007). This dry episode was even more extreme than aridity during the LGM. The Lake Malawi had lost up to 95% of its water volume and changes in a pollen structure suggest semi-arid habitats in the vicinity of the Lake (Cohen et al. 2007).

Even the post LGM wet-dry cycles were remarkable and could influence very recent intraspecific genetic division. For example, 15 000 to 5 000 years ago the Sahara Desert was nearly completely vegetated, supported numerous perennial lakes and abundant fauna (de Menocal et al. 2000).

These past climate cycles (e.g. periods of extreme climate variability, onset of lakes, LIG and LGM) led to key evolutionary changes, speciation and dispersal events (due to habitat

fragmentation) in East African mammals. Many have argued that major developments in mammalian, including human, evolution during Miocene and Plio-Pleistocene were timed with aforementioned sudden changes in Earth's climate (Trauth et al. 2005, 2007). Nevertheless, as data are stacking, this have been questioned in large mammals (Bibi and Kiessling 2015) and in smaller species as well (Montgelard and Matthee 2012). Debates about gradual or pulsed species turnover will have paramount importance for description of early human evolution (see Maslin et al. 2015 for a recent overview of the topic), and more data on different taxa is currently the only option to resolve key uncertainties currently shrouding the origin of modern humans.

1.5. Rivers, lakes, and drainage evolution

Drainage pattern (Figure 3) and historical changes in river courses undoubtedly deserve its own chapter due to the importance for phylogeographical patterns among large mammals (see Lorenzen et al. 2012 for an example on ungulates) and also small mammals (see compendium by Monadjem et al. 2015).

In particular, the Zambezi River has experienced significant changes in its course in the last five million years. The current course has been established only in the last million year (1.1–0.65 Mya; Stankiewicz and de Wit 2006, Moore et al. 2012). Models of drainage evolution in this region indicate that the Paleo-Chambeshi River flowed from its headwaters in Tanzania to the southwest into the Kalahari basin up until the Early Pleistocene and ultimately formed the current modern course of the Kafue River (Cotterill 2003b, Moore et al. 2012).

The Luangwa River hosts a specific habitat within its corridor creating conditions for phylogenetically unique taxonomical units in several species, with Thornicroft's giraffe (*Giraffa camelopardalis*

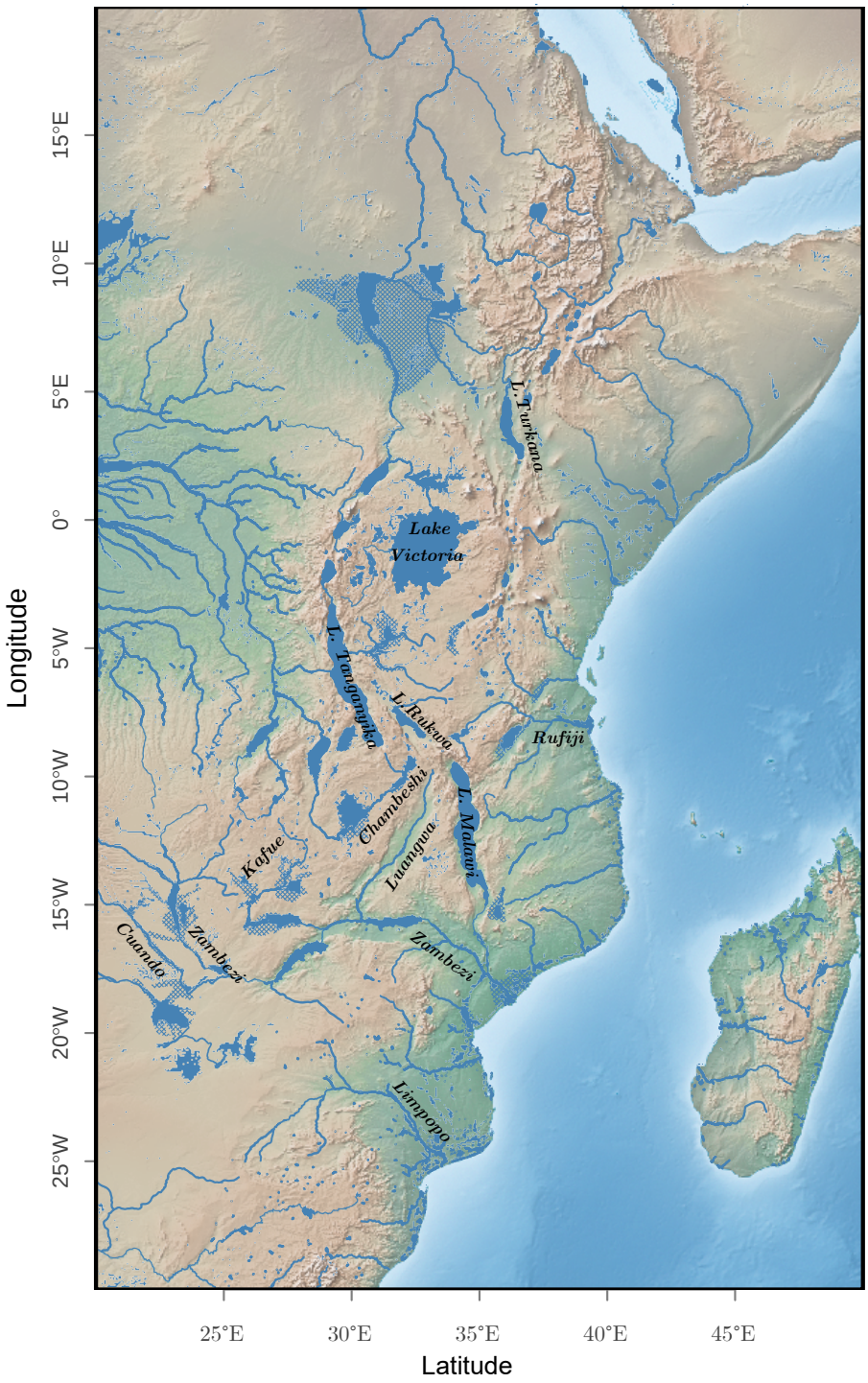


Figure 3: Hydrological map of East Africa. Lakes, rivers, and wetlands (Lehner and Döll 2004) are displayed together with habitats (grid pattern) containng phrase "flooded" in their label (Olson 2001).

thornicrofti) being a prime example (Fennessy et al. 2013). This observation was confirmed in other species such as bushbuck (Moodley and Bruford 2007). The Luangwa valley, mainly covered by dry mopane woodland, can serve as a barrier to gene flow, but as a corridor connecting South and central East Africa for arid-adapted species (Smit et al. 2011).

The great lakes, other dominant feature in Eastern Africa and possible barriers for gene flow (see Figure 3), appeared relatively late. Only Lake Malawi (Lake Nyasa) already existed in the current extent in late Pliocene, but for example Lake Victoria emerged very recently i.e. 30 000 years ago (O'Brien and Peters 1999). The historical fluctuation of the rift lakes significantly contributed to the exceptional sensitivity of East Africa to climate change compared to elsewhere on the African continent (Trauth et al. 2005, 2007).

1.6. Vegetation history of East Africa

The effects of historical habitat availability on the presence and abundance of small mammals is an important topic in understanding processes leading to current distribution of taxonomic units. Structure of vegetation is directly connected to geomorphology, climatic oscillations, and water availability, and it closely reflects global changes in these phenomena. According to Terrestrial Ecoregions of the World catalogue (Olson et al. 2001) several biomes can be found within East Africa (part of the Afrotropic realm) with wide variety of ecoregions (Figure 4).

The modern vegetation of East Africa is a complex mosaic of various types of forests, small islands of alpine vegetation including montane grasslands, thickets, vast tracts of savanna ecosystems, deserts and semi-deserts (Linder 2017); and agricultural vegetation (Mayaux et al. 2004), where many small mammals find permanent shelters and/or temporal food source. It is important to mention that

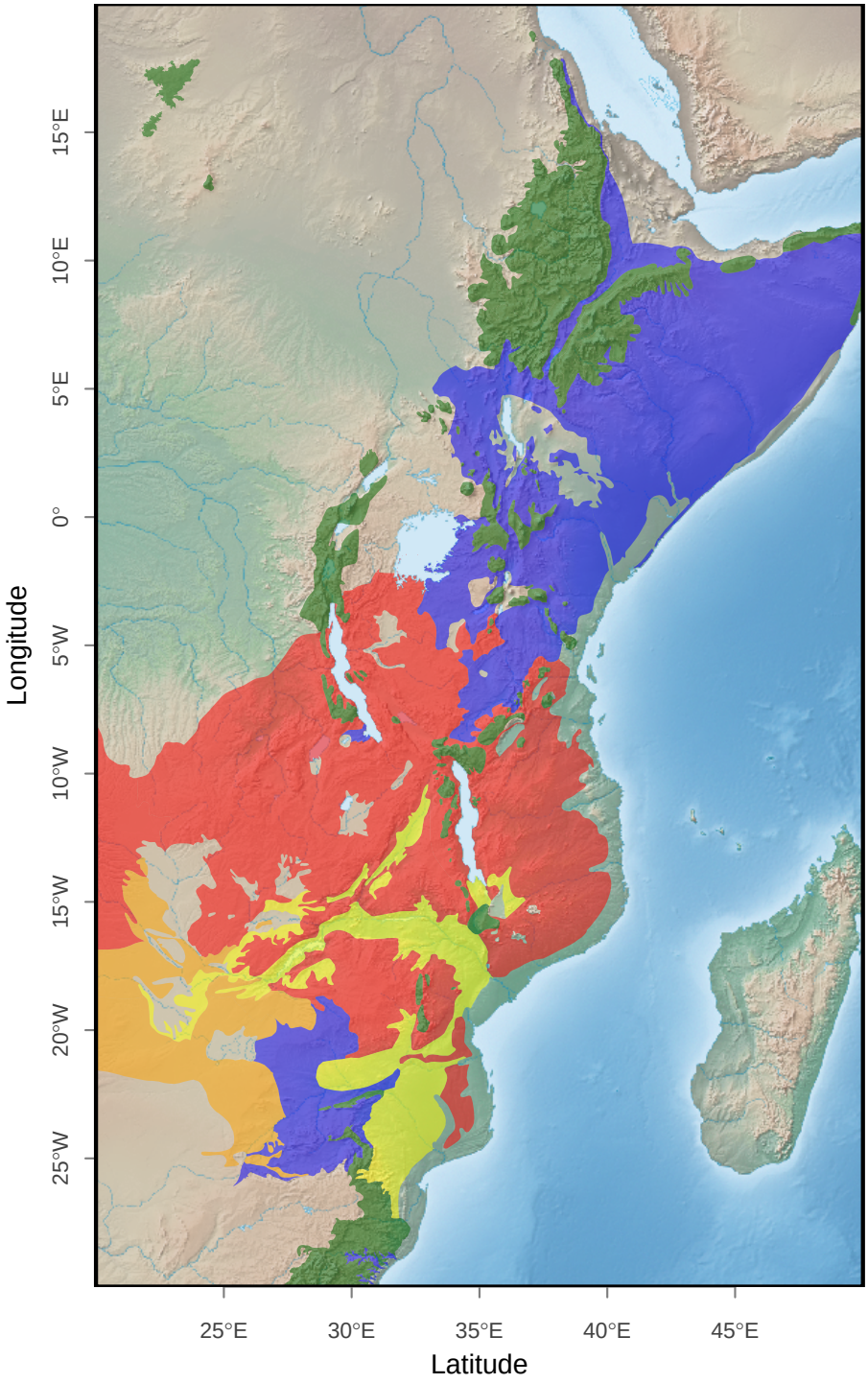


Figure 4 Savannah habitas in East Africa - Miombo woodlands (red), Mopane woodlands (yellow), Baikiaea woodlands (orange), Bushlands and thickets (blue) and Montane areas (green). Shapefiles were obtained from Terrestrial Ecoregions of the World catalogue (Olson et al. 2001).

ecoregion boundaries rarely form abrupt edges, but rather ecotones, and these mosaic habitats bound ecoregions together.

One of the most important but among the least understood terrestrial ecosystems are savannahs, primarily grassland ecosystems characterized by the trees being sufficiently widely spaced so that the canopy is not closed (Ratnam et al. 2011). Savannah habitats cover half of the African and approximately 20% of the Earth's land surface (Kahiu and Hanan 2018). The open canopy allows sufficient light to reach the ground to support an unbroken herbaceous layer. A characteristic feature is seasonal water availability, with the majority of rainfall confined to one season. Different types of savannahs (including savannah-woodland mosaic) are generally defined by the annual rainfall and tree density (Sankaran et al. 2005). This division is intricate in Eastern Africa, where the EARS makes the pattern of interconnected habitats even more complex (see Figure 4).

Of the savannas in the southern hemisphere, by far the most common is *Brachystegia* dominated miombo woodlands, the main element of the Zambezian phytochorion (White 1983). Due to nutrient-poor soils the biomass of large mammals is low in miombo (only around 20% of the grazer biomass expected in savanna) and dominated by large-bodied species (Frost and Campbell 1996). However, rodent communities in savannahs are highly variable in species composition although resource limited as well (Linzey and Kesner 1997).

The existing habitat structure is obviously an important factor forming present distributional pattern of African mammals, but understanding the dynamics of East African paleovegetation is crucial for interpreting the setting in which contemporary evolution events came to pass. Mid-Holocene and LGM climates are often equivalent if not better predictors of community structure of African mammals than modern climate (Rowan et al. 2016). In general,

during the coldest historical periods, sub-Saharan Africa was more arid and desert areas expanded, whereas forest habitats were severely constricted. During wetter periods, forest habitats replaced significant part of savannah (de Menocal 1995, 2004), resulting in gene flow restriction among isolated populations of savannah species.

For example, plateau in central Tanzania is surrounded by a "Montane forest circle" formed by southern and western Rift extended with Eastern Arc Mountains (Taylor et al. 2009). Forest biomes of East Africa might be periodically appearing and disappearing under fluctuating environmental conditions during wet Pleistocene phases (Fer et al. 2016). As suggested by locations of the contact zones between taxonomic units in multimammate mouse *Mastomys natalensis* (Colangelo et al. 2013), wet phases could have created a savanna-woodland plateau encircled by the forest biome, a habitat virtually impenetrable for non-forest species.

1.7. Small mammals as models for phylogeographic studies

The above-mentioned studies showed relatively low genetic differentiation in large mammals (Barnett et al. 2014, Bertola et al. 2016, Dubach et al. 2005, Fennessy et al. 2016, Haus et al. 2013, Lorenzen et al. 2012, Pedersen et al. 2018, Smitz et al. 2013, Zinner et al. 2013) possibly as a result of high historical and/or ongoing gene flow linked with species characteristics such as the ability to survive in a wide range of habitats and the potential for long distance dispersal and migration.

In contrast, small mammals represent suitable candidates for the reconstruction of the biotic history on a finer scale due to short generation times, limited dispersal ability and strong associations with particular habitats. For instance, periodic fragmentation of

African forests during the Plio-Pleistocene caused by paleoclimatic changes resulted in remarkable genetic differentiation of forest dwelling small non-volant mammals (e.g. Bohoussou et al. 2015, Bryja et al. 2017, 2018, Lavrenchenko and Verheyen 2005, Nicolas et al. 2008).

Phylogeographic studies on small mammals in tropical Africa focused on forests isolated on mountain ranges (e.g. Bryja, Mikula et al. 2014a, Nicolas et al. 2008) revealed strong phylogeographic structure, complex biogeographical history including dispersal–vicariance events, and brought interesting insight into the role of climatic oscillations and periodic fragmentation of forest habitats.

We decided to shift our attention to more open habitats as detailed phylogeographic analyses of taxa potentially useful for reconstructions of the evolution of these habitats in East Africa are much scarcer, although several studies in the Sudanian savanna have provided insight into the evolutionary processes that could have been responsible for the current geographic distributions of West African taxa (Brouat et al. 2009, Bryja et al. 2010, Granjon et al. 2012, Dobigny et al. 2013, Nicolas et al. 2009).

It is clear, that processes of rifting and uplifting of the Rift valley caused a mosaic of isolation and allopatric speciation events in different periods. For instance, the geomorphological features of EARS have been shown to play a prominent roles as barriers to gene flow not only in large mammals (Arctander et al. 1999, Dubach et al. 2005), but also in rodent taxa living in non-forest habitats such as savannah-dwelling *Aethomys* (Russo et al. 2006) and *Micaelamys* (Russo et al. 2010), synanthropic multimammate mouse (Colangelo et al. 2013), subterranean mammals (Faulkes et al. 2011, van Daele et al. 2007, Šumbera et al. 2018), and species adapted to semi-arid environments like spiny mice (genus *Acomys*; Verheyen et al. 2011, Petružela et al. 2018). Phylogenetic analyses of selected species can

be used to infer scenarios on various geographic levels. For example, wide distribution of the multimammate mouse allows testing hypotheses about the geomorphological and climatic history of the whole continent (Colangelo et al. 2013), and its high abundance allows to observe extremely recent splits (< 150 years) between suburban and rural populations (Gryseels et al. 2016).

Rukwa Rift and the Eastern Arc Mountains forms an important geomorphological feature responsible for separation of phylogenetic lineages in the several rodent taxa living in non-forest habitats such as the multimammate mouse (Colangelo et al. 2013, Gryseels et al. 2017), the silvery mole-rat (*Heliophobius argenteocinereus*, Faulkes et al. 2011), and the spiny mice (*Acomys spinosissimus* complex, Verheyen et al. 2011). In each of these rodent species there are phylogeographic lineages separated either by the Eastern Arc Mountains, the Rukwa Rift or both. The same pattern was observed in the bushbuck (*Tragelaphus scriptus*, Moodley and Bruford 2007). Rifting of the Tanganyika-Rukwa-Nyasa advanced in three major pulses: Miocene (8.6-5.4 Mya), Pliocene (2.4-1.7 Mya), and mid-Pleistocene to Holocene (0.57 Mya-recent) (Cotterill 2003b). These timeframes broadly coincide with major lineage splits in many above-mentioned species.

Importance of rivers and climatic fluctuations in the divergence of genealogical lineages were well demonstrated in rodents inhabiting the belt of Sudanian savannah (e.g. Brouat et al. 2009, Bryja et al. 2010, Dobigny et al. 2013). Drainage evolution of the Zambezi and its tributaries, helped to create a vicariant pattern in populations of both small and large mammals. Especially the Kafue Flats, an area of vast periodical wetland around the Lower Kafue River (Zambia), forms a transitional zone among lineages in several species of non-volant rodents. For example, a contact zone between two subspecies of pouched mice (*Saccostomus*, Mikula et al. 2016) is presented in

the Kafue Flats and Middle to Lower Zambezi. In subterranean mole-rats of the genus *Fukomys*, speciation processes took place in areas with a particularly affected geomorphology such as the capture elbows of the Zambezi, Kafue, and Chambeshi rivers (Van Daele et al. 2004, van Daele et al. 2007). For the large mammals, Kafue Flats are part of a biogeographical divider in the distribution of baboons (Papio, Zinner et al. 2009), giraffes (Fennessy et al. 2016), and reduncine antelopes (Cotterill 2003a, 2003c).

Different vegetation types could be major driving factor responsible for the current intraspecific genetic variability in southern Africa (Russo et al. 2010, du Toit et al. 2012). Recent studies from the eastern African savannah-woodland mosaic illustrated the processes involved in the forming of the fauna of open habitats and the effects of natural barriers (rivers and mountains) in combination with habitat specificity (Aghová et al. 2017, Colangelo et al. 2013, Nicolas et al. 2008, Mikula et al. 2016).

1.8. The practical importance of understanding diversity and phylogeography of East African rodents

Among small mammals, rodents are by far the largest group with over 2200 species, forming 42 % of mammalian species diversity (Wilson and Reeder 2005). Shifts in small mammal communities can alter vegetative structure and thus have implications for savannah ecosystems (Hurst and Monadjem 2014). Rodents can affect subtropical woodland-savannah habitats in many aspects. They are important part of food chain: as consumers they allow for increased plant diversity and forest regeneration, and as prey they support a diversity of carnivorous species (Linzey and Kesner 1997). Altogether, composition and density of small mammal assemblages

can serve as indicator of environmental quality. Although their importance for ecosystem is crucial, rodents as agricultural pests can inflict considerable economic damage (Singleton et al. 2010) and can be problematic in terms public health as well.

Small mammals in Africa serve as important hosts of arenaviruses, primarily rodent-borne RNA viruses causing viral hemorrhagic fevers. Although East Africa is not in the high-risk zone for pathogenic arenaviruses like Lassa (Ogbu et al. 2007), the outbreak of LUJO, arenavirus with 80% fatal rate, in Zambia demonstrated that pathogenic arenaviruses could be more widely prevalent in Africa than expected (Briese et al. 2009, Paweska et al. 2009).

Both these aspects (agriculture and public health issues) require field studies conducted on an appropriate scale. Utilization of phylogeography and landscape genetics promise facilitation of our understanding of how geographical and environmental features structure genetic variation of rodent hosts on population and individual levels (Gryseels et al. 2017). The obtained knowledge can similarly assist in developing rodent control strategies that have less reliance on expensive chemical rodenticides (Manel et al. 2010).

One of the most discussed topics in biological sciences is the origin of modern humans. Proposed in the 19th century by Charles Darwin (Darwin 1871), East Africa holds since then a central role as the environment for evolution of humankind (Bobe et al. 2002, Bobe and Behrensmeyer 2004, Bonnefille 2010, Maslin et al. 2014, Potts 1998, 2013, Reed 1997). Paleontological and archaeological evidence strongly implies that open habitats in Eastern Africa were a source of the key speciation and dispersal events in mammals (Gibernau and Montuire 1996, Potts 1998), as well as many of the critical transitions in human evolution, including the origin of the hominin clade, each genus within the clade, the extant human species, and key adaptations including bipedalism, and other behavioral innovations

(Potts 2013, Maslin et al. 2015). African rodents, were undergoing evolutionary changes in the similar environment and in the similar time frame (Fabre et al. 2012, Lecompte et al. 2008, Montgelard and Matthee 2012).

Thus, understanding of evolutionary history of small mammals in the region, the importance of geomorphologic features as barriers to gene flow, the role of frequent and abrupt habitat changes, as well as the timing of lineage splits, could bring relevant bits of knowledge required to understand the context in which early humans evolved.

2. Aims and scope of the thesis

Articles presented in this thesis focused on species of rodents inhabiting savanna-woodland biomes of the East African sub-region.

Main aims were to:

1) complete sampling in the problematically accessible areas, as detailed geographical sampling is necessary for obtaining the correct biogeographical scenarios;
2) assess the genetic diversity within the selected taxa using mitochondrial and/or nuclear genes;
3) reveal hidden diversity, cryptic species, and identify phylogenetic position of defined taxonomical units; and
4) determine approximate dates of divergences of the main lineages and possible occurrences of demographic expansions.

Following proper knowledge about genetic diversity, phylogenetic relationships, and the time frame of the origin of the taxonomical units, it would be possible to put these findings into the historical perspective and

5) identify geomorphological factors and/or climatic processes responsible for the current distribution of identified taxonomical units, with close inspection of physiographic barriers, fluvial dynamics, and climate fluctuation, i.e. factors closely linked with changes in habitat structure.

3. References

Aghová, T, R Šumbera, L Piálek, O Mikula, MM McDonough, LA Lavrenchenko, Y Meheretu, JS Mbau, and J Bryja. 2017. Multilocus phylogeny of East African gerbils (Rodentia, *Gerbilliscus*) illuminates the history of the Somali-Masai savanna. *Journal of Biogeography* 44(10) 2295–307.

Arctander, P, C Johansen, and MA Coutellec-Vreto. 1999. Phylogeography of three closely related African bovids (tribe Alcelaphini). *Molecular Biology and Evolution* 16(12) 1724–39.

Baker, RJ, and RD Bradley. 2006. Speciation in Mammals and the Genetic Species Concept. *Journal of Mammalogy* 87(4) 643–62.

Barnett, R, N Yamaguchi, B Shapiro, SYW Ho, I Barnes, R Sabin, L Werdelin, J Cuisin, and G Larson. 2014. Revealing the maternal demographic history of *Panthera leo* using ancient DNA and a spatially explicit genealogical analysis. *BMC Evolutionary Biology* 14 70.

Bertola, LD, H Jongbloed, KJ Van Der Gaag, P De Knijff, N Yamaguchi, H Hooghiemstra, H Bauer, P Henschel, PA White, CA Driscoll, T Tende, U Ottosson, Y Saidu, K Vrieling, and HH De Iongh. 2016. Phylogeographic Patterns in Africa and High Resolution Delineation of Genetic Clades in the Lion (*Panthera leo*). *Scientific Reports* 6 30807.

Bibi, F, and W Kiessling. 2015. Continuous evolutionary change in Plio-Pleistocene mammals of eastern Africa. *Proceedings of the National Academy of Sciences* 112(34) 10623–28.

Bobe, R, and AK Behrensmeyer. 2004. The expansion of grassland ecosystems in Africa in relation to mammalian evolution and the origin of the genus *Homo*. *Palaeogeography, Palaeoclimatology, Palaeoecology* 207(3–4) 399–420.

Bobe, R, AK Behrensmeyer, and RE Chapman. 2002. Faunal change, environmental variability and late Pliocene hominin evolution. *Journal of Human Evolution* 42(4) 475–97.

Bohoussou, KH, R Cornette, B Akpatou, M Colyn, JC Kerbis Peterhans, J Kennis, R Šumbera, E Verheyen, E N'Goran, P Katuala, and V Nicolas. 2015. The phylogeography of the rodent genus *Malacomys* suggests multiple Afrotropical Pleistocene lowland forest refugia. *Journal of Biogeography* 42(11) 2049–61.

Bonnefille, R. 2010. Cenozoic vegetation, climate changes and hominid evolution in tropical Africa. *Global and Planetary Change* 72(4) 390–411.

Briese, T, JT Paweska, LK McMullan, SK Hutchison, C Street, G Palacios, ML Khristova, J Weyer, R Swanepoel, M Egholm, ST Nichol, and WI Lipkin. 2009. Genetic detection and characterization of Lujo virus, a new hemorrhagic fever-associated arenavirus from southern Africa. *PLoS Pathogens* 5(5) e1000455.

Brouat, C, C Tatard, K Bâ, JF Cosson, G Dobigny, E Fichet-Calvet, L Granjon, É Lecompte, A Loiseau, K Mouline, S Piry, and JM Duplantier. 2009. Phylogeography of the Guinea multimammate mouse (*Mastomys erythroleucus*): A case study for Sahelian species in West Africa. *Journal of Biogeography* 36(12) 2237–50.

Bryja, J, L Granjon, G Dobigny, H Patzenhauerová, A Konečný, JM Duplantier, P Gauthier, M Colyn, L Durnez, A Lalis, and V Nicolas. 2010. Plio-Pleistocene history of West African Sudanian savanna and the phylogeography of the *Praomys daltoni* complex (Rodentia): The environment/geography/genetic interplay. *Molecular Ecology* 19(21) 4783–99.

Bryja, J, D Kostin, Y Meheretu, R Šumbera, A Bryjová, M Kasso, O Mikula, and LA Lavrenchenko. 2018. Reticulate Pleistocene evolution of Ethiopian rodent genus along remarkable altitudinal gradient. *Molecular Phylogenetics and Evolution* 118 75–87.

Bryja, J, O Mikula, H Patzenhauerová, NO Oguge, R Šumbera, and E Verheyen. 2014a. The role of dispersal and vicariance in the Pleistocene history of an East African mountain rodent, *Praomys delectorum*. *Journal of Biogeography* 41 196–208.

Bryja, J, O Mikula, R Šumbera, Y Meheretu, T Aghová, LA Lavrenchenko, V Mazoch, N Oguge, JS Mbau, K Welegerima, N Amundala, M Colyn, H Leirs, and E Verheyen. 2014b. Pan-African phylogeny of *Mus* (subgenus *Nannomys*) reveals one of the most successful mammal radiations in Africa. *BMC Evolutionary Biology* 14 256.

Bryja, J, R Šumbera, JC Kerbis Peterhans, T Aghová, A Bryjová, O Mikula, V Nicolas, C Denys, and E Verheyen. 2017. Evolutionary history of the thicket rats (genus *Grammomys*) mirrors the evolution of African forests since late Miocene. *Journal of Biogeography* 44(1) 182–94.

Burgin, CJ, JP Colella, PL Kahn, and NS Upham. 2018. How many species of mammals are there? *Journal of Mammalogy* 99(1) 1–14.

Chorowicz, J. 2005. The East African rift system. *Journal of African Earth Sciences* 43(1–3) 379–410.

Cohen, AS, JR Stone, KRM Beuning, LE Park, PN Reinthal, D Dettman, CA Scholz, TC Johnson, JW King, MR Talbot, ET Brown, and SJ Ivory.

2007. Ecological consequences of early Late Pleistocene megadroughts in tropical Africa. *Proceedings of the National Academy of Sciences of the United States of America* 104(42) 16422–27.

Colangelo, P, E Verheyen, H Leirs, C Tatard, C Denys, G Dobigny, JM Duplantier, C Brouat, L Granjon, and É Lecompte. 2013. A mitochondrial phylogeographic scenario for the most widespread African rodent, *Mastomys natalensis*. *Biological Journal of the Linnean Society* 108(4) 901–16.

Cotterill, FPD. 2003a. A biogeographic review of tsessebe antelopes, *Damaliscus lunatus* (Bovidae: Alcelaphini), in south-central Africa. *Durban Museum Novitates* 28 45–55.

Cotterill, FPD. 2003b. Geomorphological influences on vicariant evolution in some African mammals in the Zambezi Basin: some lessons for conservation. In *Proceedings of the Ecology and Conservation of Mini-Antelope: An International Symposium on Duiker and Dwarf Antelope in Africa*, edited by A. Plowman, 11–58. Filander Verlag.

Cotterill, FPD. 2003c. Insights into the taxonomy of tsessebe antelopes, *Damaliscus lunatus* (Bovidae: Alcelaphini) in south-central Africa: with the description of a new evolutionary species. *Durban Museum Novitates* 29 11–30.

Daele, PAAG. Van, P Dammann, JL Meier, M Kawalika, C Van De Woestijne, and H Burda. 2004. Chromosomal diversity in mole-rats of the genus *Cryptomys* (Rodentia: Bathyergidae) from the Zambezian region: With descriptions of new karyotypes. *Journal of Zoology* 264(3) 317–26.

Daele, PAAG van, E Verheyen, M Brunain, and D Adriaens. 2007. Cytochrome b sequence analysis reveals differential molecular evolution in African mole-rats of the chromosomally hyperdiverse genus *Fukomys* (Bathyergidae, Rodentia) from the Zambezian region. *Molecular Phylogenetics and Evolution* 45(1) 142–57.

Darwin, C. 1871. *The Descent of Man, and Selection in Relation to Sex*. *Botanical Review*. Vol. 11.

Demos, TC, JC Kerbis Peterhans, B Agwanda, and MJ Hickerson. 2014. Uncovering cryptic diversity and refugial persistence among small mammal lineages across the Eastern Afromontane biodiversity hotspot. *Molecular Phylogenetics and Evolution*. 71(2) 41–54.

Dobigny, G, C Tatard, P Gauthier, K Ba, J-M Duplantier, L Granjon, and GJ Kergoat. 2013. Mitochondrial and nuclear genes-based phylogeography of *Arvicanthis niloticus* (Murinae) and sub-Saharan

open habitats pleistocene history. *PloS One* 8(11) e77815.

Dubach, J, et al., B Patterson, M Briggs, K Venzke, J Flamand, P Stander, L Scheepers, and R Kays. 2005. Molecular genetic variation across the southern and eastern geographic ranges of the African lion, *Panthera leo*. *Conservation Genetics* 6(1) 15-24.

Fabre, P-H, L Hautier, D Dimitrov, P Douzery, and J Emmanuel. 2012. A glimpse on the pattern of rodent diversification: a phylogenetic approach. *BMC Evolutionary Biology* 12(1) 88.

Faulkes, CG, NC Bennett, FPD Cotterill, W Stanley, GF Mgode, and E Verheyen. 2011. Phylogeography and cryptic diversity of the solitary-dwelling silvery mole-rat, genus *Heliophobius* (family: Bathyergidae). *Journal of Zoology* 285(4) 324-38.

Feakins, SJ, NE Levin, HM Liddy, A Sieracki, TI Eglinton, and R Bonnefille. 2013. Northeast african vegetation change over 12 m.y. *Geology* 41(3) 295-98.

Fennessy, J, T Bidon, F Reuss, M Vamberger, U Fritz, V Kumar, P Elkan, MA Nilsson, and A Janke. 2016. Multi-locus analyses reveal four giraffe species instead of one. *Current Biology* 26(18) 2543-49.

Fennessy, J, F Bock, A Tutchings, R Brenneman, and A Janke. 2013. Mitochondrial DNA analyses show that Zambia's South Luangwa Valley giraffe (*Giraffa camelopardalis thornicrofti*) are genetically isolated. *African Journal of Ecology* 51(4) 635-40.

Fer, I, B Tietjen, F Jeltsch, and M H. 2016. Modelling vegetation change during Late Cenozoic uplift of the East African plateaus. *Palaeogeography, Palaeoclimatology, Palaeoecology*. Elsevier B.V.

Fick, SE, and RJ Hijmans. 2017. WorldClim 2: new 1-km spatial resolution climate surfaces for global land areas. *International Journal of Climatology* 37(12). John Wiley & Sons, Ltd 4302-15.

Frost, P, and B Campbell. 1996. The Ecology of Miombo Woodlands. *The Miombo in Transition: Woodlands and Welfare in Africa*. Bogor, Indonesia: Center for International Forestry Research: Bogor, Indonesia, 266.

Gibernau, M, and S Montuire. 1996. Mammal diversity and environment evolution during the Plio-Pleistocene in East Africa. *Human Evolution* 11(3-4) 193-204.

Gornitz, V. (ed.) 2008. Encyclopedia of Paleoclimatology and Ancient Environments. *Springer Science and Business Media*.

Granjon, L, P Colangelo, C Tatard, M Colyn, G Dobigny, and V Nicolas. 2012. Intrageneric relationships within *Gerbilliscus* (Rodentia, Muridae, Gerbillinae), with characterization of an additional West African species. *Zootaxa* 3325 1-25.

Groves, CP. 2013. The nature of species: A rejoinder to Zachos et al. *Mammalian Biology* 78 7-9.

Groves, CP, FPD Cotterill, S Gippoliti, J Robovský, C Roos, PJ Taylor, and D Zinner. 2017. Species definitions and conservation: a review and case studies from African mammals. *Conservation Genetics* 18(6). Springer Netherlands 1247-56.

Gryseels, S, SJE Baird, B Borremans, R Makundi, H Leirs, JG de Bellocq, and JG de Bellocq. 2017. Viruses don't go viral: the importance of host phylogeographic structure in the spatial spread of arenaviruses. *PLoS Pathogens* 13(1) e1006073.

Gryseels, S, JG de Bellocq, R Makundi, K Vanmechelen, J Broeckhove, V Mazoch, R Šumbera, J Zima, H Leirs, and SJE Baird. 2016. Genetic distinction between contiguous urban and rural multimammate mice in Tanzania despite gene flow. *Journal of Evolutionary Biology* 29(10) 1952-67.

Haus, T, E Akom, B Agwanda, M Hofreiter, C Roos, and D Zinner. 2013. Mitochondrial Diversity and Distribution of African Green Monkeys (*Chlorocebus* Gray, 1870). *American Journal of Primatology* 75(1) 350-60.

Hurst, ZM, and A Monadjem. 2014. Linking changes in small mammal communities to ecosystem functions in an agricultural landscape. *Mammalian Biology - Zeitschrift Für Säugetierkunde* 79(1) 17-23.

Imbrie, J, EA Boyle, SC Clemens, A Duffy, WR Howard, G Kukla, J Kutzbach, DG Martinson, A McIntyre, AC Mix, B Molfino, JJ Morley, LC Peterson, NG Pisias, WL Prell, ME Raymo, NJ Shackleton, and JR Toggweiler. 1992. On the Structure and Origin of Major Glaciation Cycles 1. Linear Responses to Milankovitch Forcing. *Paleoceanography* 7(6) 701-38.

Kahiu, MN, and NP Hanan. 2018. Estimation of Woody and Herbaceous Leaf Area Index in Sub-Saharan Africa Using MODIS Data. *Journal of Geophysical Research: Biogeosciences* 123(1) 3-17.

Lavrenchenko, LA, and E Verheyen. 2005. An assessment of the systematics of the genus Desmomys Thomas, 1910 (Rodentia: Muridae) using mitochondrial DNA sequences. *African Biodiversity* 1910 363-69.

Lecompte, É, K Aplin, C Denys, F Catzeflis, M Chades, and P Chevret. 2008. Phylogeny and biogeography of African Murinae based on mitochondrial and nuclear gene sequences, with a new tribal classification of the subfamily. *BMC Evolutionary Biology* 8(1) 199.

Linder, HP. 2017. East African Cenozoic vegetation history. *Evolutionary Anthropology* 26(6) 300-312.

Linzey, AV., and MH Kesner. 1997. Small mammals of a woodland-savannah ecosystem in Zimbabwe. I. Density and habitat occupancy patterns. *Journal of Zoology* 243(1) 137-152.

Lehner, B. and Döll, P. (2004). Development and validation of a global database of lakes, reservoirs and wetlands. *Journal of Hydrology* 296(1-4) 1-22.

Lorenzen, ED, R Heller, and HR Siegismund. 2012. Comparative phylogeography of African savannah ungulates. *Molecular Ecology* 21(15) 3656-70.

Macgregor, D. 2015. History of the Development of the East African Rift System: A Series of Interpreted Maps through Time. *Journal of African Earth Sciences* 101 232-252.

Manel, S, S Joost, BK Epperson, R Holderegger, A Storfer, MS Rosenberg, KT Scribner, A Bonin, and MJ Fortin. 2010. Perspectives on the use of landscape genetics to detect genetic adaptive variation in the field. *Molecular Ecology* 19(17) 3760-72.

Maslin, MA, CM Brierley, AM Milner, S Shultz, MH Trauth, and KE Wilson. 2014. East african climate pulses and early human evolution. *Quaternary Science Reviews* 101 1-17.

Maslin, MA, S Shultz, and MH Trauth. 2015. A synthesis of the theories and concepts of early human evolution. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences* 370(1663) 1-12.

Mayaux, P, E Bartholomé, S Fritz, and A Belward. 2004. A new land-cover map of Africa for the year 2000. *Journal of Biogeography* 31(6) 861-77.

Menocal, PB de. 1995. Plio-Pleistocene African Climate. *Science* 270(5233) 53-59.

Menocal, PB de. 2004. African climate change and faunal evolution during the Pliocene-Pleistocene. *Earth and Planetary Science Letters* 220(1-2) 3-24.

Menocal, PB de, J Ortiz, T Guilderson, J Adkins, M Sarnthein, L Baker,

and M Yarusinsky. 2000. Abrupt onset and termination of the African Humid Period: Rapid climate responses to gradual insolation forcing. *Quaternary Science Reviews* 19(1-5) 347-61.

Mikula, O, R Šumbera, T Aghová, JS Mbau, AS Katakweba, CA Sabuni, and J Bryja. 2016. Evolutionary history and species diversity of African pouched mice (Rodentia: Nesomyidae: *Saccostomus*). *Zoologica Scripta* 45(6) 595-617.

Monadjem, A, PJ Taylor, C Denys, and FPD Cotterill. 2015. *Rodents of Sub-Saharan Africa: A Biogeographic and Taxonomic Synthesis*. Walter de Gruyter GmbH & Co KG.

Montgelard, C, and CA Matthee. 2012. Tempo of genetic diversification in southern African rodents: The role of Plio-Pleistocene climatic oscillations as drivers for speciation. *Acta Oecologica* 42 50-57.

Moodley, Y, and MW Bruford. 2007. Molecular biogeography: Towards an integrated framework for conserving Pan-African biodiversity. *PLoS ONE* 2(5) e454.

Moore, AE, FPD Cotterill, and FD Eckardt. 2012. The evolution and ages of Makgadikgadi palaeo-lakes: Consilient evidence from Kalahari drainage evolution South-Central Africa. *South African Journal of Geology* 115(3) 385-413.

Nicolas, V, J Bryja, B Akpatou, A Konecny, É Lecompte, M Colyn, A Lalis, A Couloux, C Denys, and L Granjon. 2008. Comparative phylogeography of two sibling species of forest-dwelling rodent (*Praomys rostratus* and *P. tullbergi*) in West Africa: Different reactions to past forest fragmentation. *Molecular Ecology* 17(23) 5118-34.

Nicolas, V, L Granjon, JM Duplantier, C Cruaud, and G Dobigny. 2009. Phylogeography of spiny mice (genus *Acomys*, Rodentia: Muridae) from the south-western margin of the sahara with taxonomic implications. *Biological Journal of the Linnean Society* 98(1) 29-46.

Nyblade, AA, and SW Robinson. 1994. The African Superswell. *Geophysical Research Letters* 21(9) 765-68.

O'Brien, E, and CR Peters. 1999. Landforms, climate, ecogeographic mosaics, and the potential for hominid diversity in Pliocene Africa. In *African Biogeography, Climate Change & Human Evolution*, 115-37.

Ogbu, O, E Ajuluchukwu, and C Uneke. 2007. Lassa fever in West African sub-region: an overview. *Journal of Vector Borne Diseases* 44(1) 1.

Olson, DM, E Dinerstein, ED Wikramanayake, ND Burgess, GVN Powell, EC Underwood, JA D'amico, I Itoua, HE Strand, JC Morrison, CJ

Loucks, TF Allnutt, TH Ricketts, Y Kura, JF Lamoreux, WW Wettengel, P Hedao, KR Kassem, ED Wikramanaya, ND Burgess, GVN Powell, EC Underwood, JA D'amico, I Itoua, HE Strand, JC Morrison, CJ Loucks, TF Allnutt, TH Ricketts, Y Kura, JF Lamoreux, WW Wettengel, P Hedao, and KR Kassem. 2001. Terrestrial Ecoregions of the World: A New Map of Life on Earth. *BioScience* 51(11) 933-938.

Otto-Bliesner, BL, SJ Marshall, JT Overpeck, GH Miller, and A Hu. 2006. Simulating Arctic climate warmth and icefield retreat in the last interglaciation. Science 311(5768) 1751-53.

Paweska, J, N Sewlall, T Ksiazek, L Blumberg, M Hale, W Lipkin, J Weyer, ST Nichol, P Rollin, and L McMullan. 2009. Nosocomial outbreak of novel arenavirus infection, Southern Africa. *Emerging Infectious Diseases* 15(10) 1598.

Pedersen, CET, A Albrechtsen, PD Etter, EA Johnson, L Orlando, L Chikhi, HR Siegismund, and R Heller. 2018. A southern African origin and cryptic structure in the highly mobile plains zebra. *Nature Ecology and Evolution* 2(3) 491-98.

Petružela, J, R Šumbera, T Aghová, A Bryjová, AS Katakweba, CA Sabuni, WN Chitaukali, and J Bryja. 2018. Spiny mice of the Zambezian bioregion - phylogeny, biogeography and ecological differentiation within the Acomys spinosissimus complex. *Mammalian Biology* 91 79-90.

Phillips, SJ, RP Anderson, and RE Schapire. 2006. Maximum entropy modeling of species geographic distributions. *Ecological Modelling* 190(3-4) 231-59.

Potts, R. 1998. Environmental hypotheses of hominin evolution. *American Journal of Physical Anthropology* 107(S27) 93-136.

Potts, R. 2013. Hominin evolution in settings of strong environmental variability. *Quaternary Science Reviews* 73 1-13.

R Core Team. 2018. R: A Language and Environment for Statistical Computing. Vienna, Austria.

Ratnam, J, WJ Bond, RJ Fensham, WA Hoffmann, S Archibald, CER Lehmann, MT Anderson, SI Higgins, and M Sankaran. 2011. When is a "forest" a savanna, and why does it matter? *Global Ecology and Biogeography* 20(5) 653-60.

Reed, KE. 1997. Early hominid evolution and ecological change through the African Plio-Pleistocene. *Journal of Human Evolution* 32(2-3) 289-322.

Roberts, EM, NJ Stevens, PM O'Connor, PHGM Dirks, MD Gottfried, WC Clyde, RA Armstrong, a. IS Kemp, and S Hemming. 2012. Initiation of

the western branch of the East African Rift coeval with the eastern branch. *Nature Geoscience* 5(4) 289-94.

Rowan, J, JM Kamilar, L Beaudrot, and KE Reed. 2016. Strong influence of palaeoclimate on the structure of modern African mammal communities. *Proceedings of the Royal Society B: Biological Sciences* 283(1840) 20161207.

Russo, I-RM, CT Chimimba, and P Bloomer. 2006. Mitochondrial Dna Differentiation Between Two Species of Aethomys (Rodentia: Muridae) From Southern Africa. *Journal of Mammalogy* 87(3) 545-53.

Russo, I-RM, CT Chimimba, and P Bloomer. 2010. Bioregion heterogeneity correlates with extensive mitochondrial DNA diversity in the Namaqua rock mouse, *Micaelamys namaquensis* (Rodentia: Muridae) from southern Africa - evidence for a species complex. *BMC Evolutionary Biology* 10(1) 307.

Sankaran, M, NP Hanan, RJ Scholes, J Ratnam, DJ Augustine, BS Cade, J Gignoux, SI Higgins, X Le Roux, F Ludwig, J Ardo, F Banyikwa, A Bronn, G Bucini, KK Caylor, MB Coughenour, A Diouf, W Ekaya, CJ Feral, EC February, PGH Frost, P Hiernaux, H Hrabar, KL Metzger, HHT Prins, S Ringrose, W Sea, J Tews, J Worden, and N Zambatis. 2005. Determinants of woody cover in African savannas. *Nature* 438(7069) 846-49.

Scholz, CA, TC Johnson, AS Cohen, JW King, J a Peck, JT Overpeck, MR Talbot, ET Brown, L Kalindekafe, PYO Amoako, RP Lyons, TM Shanahan, IS Castañeda, CW Heil, SL Forman, LR McHargue, KRM Beuning, J Gomez, and J Pierson. 2007. East African megadroughts between 135 and 75 thousand years ago and bearing on early-modern human origins. *Proceedings of the National Academy of Sciences of the United States of America* 104 16416-21.

Sepulchre, P, G Ramstein, F Fluteau, M Schuster, JJ Tiercelin, and M Brunet. 2006. Tectonic uplift and Eastern Africa aridification. *Science* 313(5792) 1419-23.

Singleton, GR, S Belmain, P Brown, and B Hardy. 2010. *Rodent Outbreaks: Ecology and Impacts*. *Rodent Outbreaks: Ecology and Impacts*. International Rice Research Institute.

Smit, HA, B Jansen van Vuuren, PCM O'Brien, M Ferguson-Smith, F Yang, and TJ Robinson. 2011. Phylogenetic relationships of elephant-shrews (Afrotheria, Macroscelididae). *Journal of Zoology* 284(2003) 133-43.

Smitz, N, C Berthouly, D Cornélis, R Heller, P van Hooft, P Chardonnet,

A Caron, H Prins, BJ van Vuuren, H de Iongh, and J Michaux. 2013. Pan-African Genetic Structure in the African Buffalo (*Syncerus caffer*): Investigating Intraspecific Divergence. *PLoS ONE* 8(2) e56235.

Stankiewicz, J, and MJ de Wit. 2006. A proposed drainage evolution model for Central Africa - Did the Congo flow east? *Journal of African Earth Sciences* 44(1) 75-84.

Steppan, SJ, and JJ Schenk. 2017. Muroid rodent phylogenetics: 900-species tree reveals increasing diversification rates. *PLoS ONE* 12(8) e0183070.

Šumbera, R, J Krásová, LA Lavrenchenko, S Mengistu, A Bekele, O Mikula, and J Bryja. 2018. Ethiopian highlands as a cradle of the African fossorial root-rats (genus *Tachyoryctes*), the genetic evidence. *Molecular Phylogenetics and Evolution* 126 105-15.

Taylor, PJ, S Maree, J Van Sandwyk, JC Kerbis Peterhans, WT Stanley, E Verheyen, P Kaliba, W Verheyen, P Kaleme, and NC Bennett. 2009. Speciation mirrors geomorphology and palaeoclimatic history in African laminate-toothed rats (Muridae: Otomyini) of the *Otomys denti* and *Otomys lacustris* species-complexes in the "Montane Circle" of East Africa. *Biological Journal of the Linnean Society* 96(4) 913-41.

Toit, N du, B Jansen van Vuuren, S Matthee, and CA Matthee. 2012. Biome specificity of distinct genetic lineages within the four-striped mouse *Rhabdomys pumilio* (Rodentia: Muridae) from southern Africa with implications for taxonomy. *Molecular Phylogenetics and Evolution* 65(1) 75-86.

Trauth, MH, MA Maslin, AL Deino, and M Strecker. 2005. Late Cenozoic moisture history of East Africa. *Science* 309(5743) 2051-53.

Trauth, MH, MA Maslin, AL Deino, MR Strecker, AGN Bergner, and M Dhnforth. 2007. High- and low-latitude forcing of Plio-Pleistocene East African climate and human evolution. *Journal of Human Evolution* 53(5) 475-86.

Verheyen, W, J Hulselmans, W Wendelen, H Leirs, and M Corti. 2011. Contribution to the systematics and zoogeography of the East-African *Acomys*. *Zootaxa* 3059 1-35.

White, F. 1983. The Zambezian regional centre of endemism. In *The Vegetation of Africa: A Descriptive Memoir to Accompany the Unesco/AETFAT/UNSO Vegetation Map of Africa*, edited by Frank White, 86-101. University of Michigan.

Wilson, DE, and DM Reeder. 2005. *Mammal Species of the World. A Taxonomic and Geographic Reference (3rd Ed)*. Johns Hopkins

University Press.

Wolfenden, E, C Ebinger, G Yirgu, AL Deino, and D Ayalew. 2004. Evolution of the northern Main Ethiopian rift: birth of a triple junction. *Earth and Planetary Science Letters* 224(1-2) 213-28.

Zachos, J, M Pagani, L Sloan, E Thomas, and K Billups. 2001. Trends, rhythms, and aberrations in global climate 65 Mya to present. *Science (New York, N.Y.)* 292(5517) 686-93.

Zinner, D, LF Groeneveld, C Keller, and C Roos. 2009. Mitochondrial phylogeography of baboons (*Papio* spp.): indication for introgressive hybridization? *BMC Evolutionary Biology* 9 83.

Zinner, D, J Wertheimer, R Liedigk, LF Groeneveld, and C Roos. 2013. Baboon phylogeny as inferred from complete mitochondrial genomes. *American Journal of Physical Anthropology* 150 133-40.

Chapter I

Phylogeography of a widespread sub-Saharan murid rodent *Aethomys chrysophilus*: the role of geographic barriers and paleoclimate in the Zambezian bioregion.

Mazoch V., Mikula O., Bryja J., Konvičková H., Russo I. R., Verheyen E. & Šumbera R. (2017). *Mammalia*, 82, 4, 373-387, 2018.

Vladimír Mazoch[a], Ondřej Mikula[a,*], Josef Bryja, Hana Konvičková, Isa-Rita Russo, Erik Verheyen and Radim Šumbera

Phylogeography of a widespread sub-Saharan murid rodent *Aethomys chrysophilus*: the role of geographic barriers and paleoclimate in the Zambezian bioregion

https://doi.org/10.1515/mammalia-2017-0001
Received January 7, 2017; accepted June 28, 2017

Abstract: Murid rodents of the genus *Aethomys* are one of the most common rodents in drier habitats in sub-Saharan Africa. Among them, the red veld rat *Aethomys chrysophilus* is the most widespread species with the core distribution located in the Zambezian bioregion. In this study, we describe phylogeographic structure of the species and estimate its age from a time-calibrated phylogeny of the genus. Seven parapatric clades were identified in the mitochondrial cytochrome *b* phylogeny, where some of the distributions of these clades have been separated by previously described biogeographical divides (Zambezi-Kafue river system, Rukwa Rift and the Eastern Arc Mountains). One internal clade corresponded to populations previously described as a distinct species, *Aethomys ineptus*. The whole *A. chrysophilus* complex was estimated to be 1.3 (0.5–2.4) Mya old, with *A. ineptus* originating 0.7 (0.1–1.4) Mya before present. The internal position of *A. ineptus* was also recovered in phylogenetic reconstruction based on two nuclear genes and thus it is not a consequence of mitochondrial introgression. In addition, we analyzed skull form variation across the species' distributional range and found no significant difference between *A. ineptus* and the rest of *A. chrysophilus* complex.

Keywords: *Aethomys chrysophilus*; *Aethomys ineptus*; phylogeography; Plio-Pleistocene climate changes; Zambezian bioregion.

Introduction

A recent boom in studies combining molecular genetics and geographical data provides a tool for better understanding phylogenetic diversity and historical processes responsible for the current distribution of species, especially in neglected areas across the globe. For example, in sub-Saharan Africa descriptions of genetic diversity helped to identify historical refugia and migration barriers in larger mammals, e.g. ungulates (Lorenzen et al. 2012), baboons (Zinner et al. 2013) and green monkeys (Haus et al. 2013). Despite substantial progress, we are still far from fully comprehending all factors that have affected the current distribution of biological diversity on the continent. The above-mentioned studies showed low genetic differentiation in large mammals possibly as a result of high historical and/or ongoing gene flow linked with species characteristics such as the ability to survive in a wide range of habitats and the potential for long distance dispersal. In contrast, small mammals represent suitable candidates for the reconstruction of the biotic history on a finer scale due to short generation times, limited dispersal ability and strong associations with particular habitats. For instance, periodic fragmentation of African forests during the Plio-Pleistocene caused by palaeoclimatic changes resulted in remarkable genetic differentiation of forest dwelling small non-flying mammals (e.g. Nicolas et al. 2011, Demos et al. 2014, Bohoussou et al. 2015).

On the other hand, factors affecting the evolution of species living in drier and more open habitats remain less known. Phylogeographic studies suggested the

[a]**Vladimír Mazoch and Ondřej Mikula:** These authors contributed equally to this article.
***Corresponding author: Ondřej Mikula,** Institute of Vertebrate Biology, Czech Academy of Sciences, Květná 8, Brno 60365, Czech Republic; and Institute of Animal Physiology and Genetics, Czech Academy of Sciences, Brno, Czech Republic, e-mail: onmikula@gmail.com
Vladimír Mazoch: Department of Zoology, Faculty of Science, University of South Bohemia, České Budějovice, Czech Republic; and Institute of Vertebrate Biology, Czech Academy of Sciences, Květná 8, Brno 60365, Czech Republic
Josef Bryja: Institute of Vertebrate Biology, Czech Academy of Sciences, Květná 8, Brno 60365, Czech Republic; and Department of Botany and Zoology, Faculty of Science, Masaryk University, Brno, Czech Republic
Hana Konvičková: Institute of Vertebrate Biology, Czech Academy of Sciences, Květná 8, Brno 60365, Czech Republic
Isa-Rita Russo: Cardiff School of Biosciences, Cardiff University, Sir Martin Evans Building, Museum Avenue, Cardiff, CF10 3AX, UK
Erik Verheyen: Royal Belgian Institute for Natural Sciences, Operational Direction Taxonomy and Phylogeny, 1000, Brussels, Belgium; and Evolutionary Ecology Group, Biology Department, University of Antwerp, 2020, Antwerp, Belgium
Radim Šumbera: Department of Zoology, Faculty of Science, University of South Bohemia, České Budějovice, Czech Republic

importance of rivers and climatic fluctuations in the divergence of genealogical lineages of rodents inhabiting the belt of Sudanian savanna (e.g. Bryja et al. 2010, Dobigny et al. 2013), while different vegetation types could be responsible for the current intraspecific genetic variability in southern Africa (Russo et al. 2010). Recent studies from the eastern African savanna-woodland mosaic illustrated the processes involved in the forming of the fauna of open habitats and the effects of natural barriers (rivers and mountains) in combination with habitat specificity (Nicolas et al. 2008, Colangelo et al. 2013, McDonough et al. 2015, Mikula et al. 2016, Aghová et al. 2017).

In this study, we analyzed the genetic structure of a murid rodent, the red veld rat *Aethomys chrysophilus* (de Winton 1896). The distribution of this species is tightly linked to the Miombo woodland and the adjacent drier habitats. This species can therefore serve as a suitable model for assessing the role of historical factors that shaped the evolution of African seasonal savanna-woodlands. Among the nine currently recognized *Aethomys* species, the red veld rat has by far the largest distribution range. This medium-sized rodent with reddish-brown pelage mixed with dark hairs and a long sparsely haired tail occurs from southern-most Kenya through Tanzania, Zambia and Zimbabwe to the KwaZulu-Natal province of South Africa, and from Namibia and southern Angola to southern Mozambique. Across its distribution range it is found in various savanna and woodland habitats, but absent from very arid regions and forests (Linzey and Chimimba 2008).

Gordon and Rautenbach (1980) found two cytotypes (2n = 44 and 50) in *Aethomys chrysophilus* and Chimimba (1998) designated 2n = 44 populations as a separate species *Aethomys ineptus* (Thomas and Wroughton 1908), given their distinct distribution in southern Africa and very specific sperm morphology (Visser and Robinson 1986, 1987). A more recent study based on mitochondrial DNA suggests that *A. ineptus* is monophyletic, but nested within *A. chrysophilus* from southern and eastern Africa, so the latter species is left paraphyletic (Russo et al. 2006). In addition, Russo et al. (2006) suggested further in-depth analysis to study genetic differentiation within *A. chrysophilus* with samples representing the entire distributional range of the species. This study also stressed the need for a taxonomic revision within the species to identify any additional cryptic taxa that may be contained within the currently described *A. chrysophilus*. Here we refer to both species as members of the *A. chrysophilus* complex.

Using a phylogeographic approach combined with species distribution modeling and morphometric analysis, we aimed to (1) to assess the mitochondrial diversity within the *Aethomys chrysophilus* complex using the cytochrome *b* (*CYTB*) gene; (2) to identify geomorphological factors and/or climatic processes responsible for the current distribution of mitochondrial lineages; (3) to date divergences of the main mitochondrial lineages and confirm internal position of *Aethomys ineptus* within the *A. chrysophilus* complex by analysis of nuclear markers; (4) to test for skull form differences between phylogeographic lineages.

Materials and methods

Sampling

We obtained and analyzed genetic information from 222 individuals documented in the Supplemental File 1 available via figshare (DOI: 10.6084/m9.figshare.4516745). Tissue samples (169) were collected by the authors or recovered from museum specimens, namely from collections of Centre de Biologie pour la Gestion des Populations (Montpellier, France), Royal Belgian Institute of Natural Sciences (Brussels, Belgium), Magyar Természettudományi Múzeum (Budapest, Hungary), Muséum National d'Histoire Naturelle (Paris, France), Naturmuseum Senckenberg (Frankfurt, Germany), Smithsonian Institution – National Museum of Natural History (Washington, DC, USA) and Texas Tech University (Lubbock, TX, USA). This material was supplemented by sequences from 23 georeferenced individuals downloaded from the African Rodentia database (http://projects.biodiversity.be/africanrodentia) and 30 individuals from GenBank (http://www.ncbi.nlm.nih.gov/genbank). In total, genetic data were collected from 108 georeferenced localities in 10 countries almost spanning the complete distribution of the *Aethomys chrysophilus* complex (see Supplemental File 1, Figure 1).

Genotyping

DNA from fresh (ethanol- or DMSO-preserved) tissues was extracted using the DNeasy tissue kit (Qiagen, Hilden, Germany) following the manufacturer's instructions. For the basic genetic characterization of all individuals we amplified the complete mitochondrial *CYTB* gene by polymerase chain reaction (PCR) using primers detailed in Table 1. For the purpose of divergence dating and confirmation of the observed mitochondrial diversity, we also sequenced the nuclear interphotoreceptor retinoid binding protein (*IRBP*) and the recombination activation protein (*RAG1*) of selected specimens (see more details in Table 1).

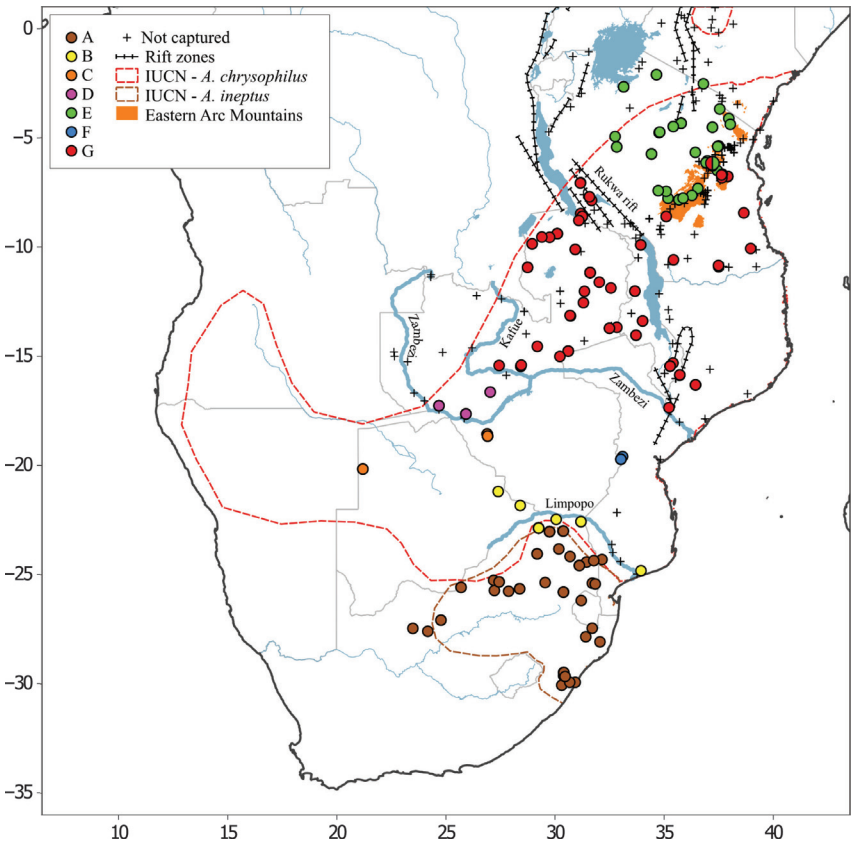


Figure 1: Geographic distribution of seven clades of the *A. chrysophilus* complex.
Trapping sites of particular clades are indicated by color that correspond to Figure 2. Black crosses represent sites where trapping was conducted but no individuals of *A. chrysophilus* complex were captured. Brown (*A. ineptus*) and red (*A. chrysophilus*) dotted lines demarcate distribution range according to IUCN. Blue color comprises lakes and main rivers. Black lines showing position of main rift faults.

Table 1: Amplified genetic markers.

Locus	Primers (5′→3′)	Ta	Reference
CYTB	L14723: ACC AAT GAC ATG AAA AAT CAT CGT T H15915: TCT CCA TTT CTG GTT TAC AAG AC	52°C/60 s	Irwin et al. 1991
IRBP	IRBP-217: ATG GCC AAG GTC CTC TTG GAT AAC TAC TGC TT IRBP-1531: CGC AGG TCC ATG ATG AGG TGC TCC GTG TCC TG	55°C/30 s	Stanhope et al. 1992
RAG1	RAG1-F1705: GCT TTG ATG GAC ATG GAA GAA GAC AT RAG1-R2951: GAG CCA TCC CTC TCA ATA ATT TCA GG	60–57°C (touch down PCR)	Teeling et al. 2000

The purified PCR products were sequenced in a commercial laboratory. Genetic data obtained from fresh material were complemented by additional sequences from 10 museum samples (mostly dry skins; see Supplemental File 1), where partial *CYTB* sequences were generated by pyrosequencing on GS Junior (Roche, Basel, Switzerland) by using the mini-barcode protocol (Galan et al. 2012). The main advantage of this approach is that it allows separating individual sequences in samples contaminated by distantly related organisms (often the case of museum samples), which is not possible through the traditional Sanger sequencing method (for more details see Bryja

et al. 2014). Obtained sequences were edited and aligned using Geneious version 9.1.5 (Biomatters, Auckland, New Zealand, available from http://www.geneious.com). Unpublished sequences used in our phylogenetic analyses were submitted to GenBank (accession numbers KY965315–KY965392, KU723654, KU723655, KU723662, KU723668, KU723672, KU747156, KU747157).

Phylogenetic reconstructions and historical demography

The *CYTB* phylogeny was inferred from 86 unique haplotypes using a Bayesian inference and a maximum likelihood (ML) approach the best-fit model of evolution (GTR+G) was selected using jModelTest 2 (Darriba et al. 2012). The Bayesian analysis in MrBayes 3.2.6 (Ronquist et al. 2012) consisted of two independent Markov chain Monte Carlo (MCMC) runs whose mixing and convergence was checked in Tracer v1.6 (Rambaut and Drummond 2013) with default priors on all parameters. A 10% burn-in was sufficient to ensure that trees were only sampled after MCMC reached its equilibrium distribution. The ML phylogeny was estimated by RA×ML 8.6.2 (Stamatakis 2014) using rapid bootstrapping (1000 replicates) to evaluate support for internal nodes (Stamatakis et al. 2008). The same three outgroups (*Aethomys hindei*, *Aethomys kaiseri*, *Aethomys nyikae*) were used in both analyses. All phylogenetic and divergence dating computations were performed on the CIPRES cluster (Miller et al. 2010). An alternative view of *CYTB* variation was provided by the neighbor-net haplotype network method in Splits Tree 4.0 (Huson and Bryant 2006).

Eighty-six haplotypes included in the *CYTB* phylogeny represented 129 individuals. Remaining shorter sequences from 82 individuals were placed *post hoc* into the majority consensus Bayesian tree by the evolutionary placement algorithm (EPA; Berger et al. 2011). The EPA accepts the tree topology and sequence alignment as inputs, performs ML estimations of GTR+G parameters and branch lengths on the fixed topology. Query sequences were then taken one by one to estimate their ML placements in the phylogeny.

We calculated Tajima's D (Tajima 1989) and Fu's F_s (Fu 1997) statistics for the two clades with more than 10 haplotypes to test whether the observed *CYTB* variation conformed to expectation for a neutrally evolving locus in mutation-drift equilibrium. In addition, predictions of a sudden expansion model were assessed using mismatch distributions and Harpending raggedness index (Rogers and Harpending 1992). Historical demography statistics were calculated using DnaSP 5.10.1 (Librado and Rozas 2009) and mismatch distributions were visualized using the MMD function as implemented in the R package "pegas" (Paradis 2010).

Dating of divergences

The oldest fossil that has been reliably identified as belonging to the genus *Aethomys* has been found with other Arvicanthini (*Arvicanthis*, *Lemniscomys*) in Kenya, Lemudong'o (Manthi 2007), estimated at 6.08–6.12 million years ago (Mya) (Deino and Ambrose 2007). The only other Arvicanthini fossil has been reported by Mein et al. (2004) from Harasib in Namibia. However, the lower molars were very untypical of *Aethomys* with some uncertainty around the date of this fossil. The faunal composition was similar to Ethiopian Chorora which may be as old as 10–11 Mya (Geraads et al. 2002) or 8.5 Mya (Suwa et al. 2015). Thus, we chose to estimate minimum divergence times by setting a monotonically decreasing calibration for the root of the *Aethomys* phylogeny. We used an exponential distribution with a mean of 1.504 and an offset of six which gives the highest probability to the minimum age of 6 Mya and only 5% of the probability density to age older than 10.5 Mya. However, the actual priors of node ages were different due to the inferred fossil calibration density with the birth-death tree shape prior (Heled and Drummond 2012) and we therefore explored this by MCMC sampling. The 95% probability density was between 6.0 and 9.4 Mya for the root and 0.0 and 5.2 Mya for the most recent common ancestor (MRCA) of *Aethomys chrysophilus* (or any other monophyletic group of three sequences).

The dating was performed in BEAST 2.3.2 (Bouckaert et al. 2014) using nuclear gene (*IRBP*, *RAG1*) sequences of nine *Aethomys* specimens (see Supplemental File 1), each representing one of the major lineages (species or intraspecific phylogroups) within the genus. More specifically, we included three lineages of the *Aethomys chrysophilus* complex including *Aethomys ineptus* to test for its internal placement based on nuclear genes only. Following preliminary analyses in jModelTest 2 we used the K80 model of nucleotide substitution (Kimura 1980). We assumed a relaxed molecular clock with uncorrelated lognormal distribution of evolutionary rates (Drummond et al. 2006) and the same birth-death tree for all genes. Two separate runs were conducted and combined after checking for mixing and convergence with a 10% burn-in. A maximum clade credibility tree was determined using the TreeAnnotator tool of BEAST. Tree figures were produced using R software (R Core Team 2016) using the package "ape" (Paradis et al. 2004).

Species distribution modeling

For the species distribution modeling, we merged all georeferenced records of the *Aethomys chrysophilus* complex from our own and published data sets (Gordon and Rautenbach 1980, Gordon and Watson 1986, Visser and Robinson 1986, 1987, Baker et al. 1988, Breed et al. 1988, Ducroz et al. 2001, Fadda et al. 2001, Castiglia et al. 2003, Linzey et al. 2003, Russo et al. 2006, Nicolas et al. 2011, Phukuntsi et al. 2016, this study) in which species identity was confirmed by mtDNA, karyotypes and sperm morphology (see Supplemental File 1). After discarding samples from identical localities and rounding latitudes and longitudes to 0.5° we obtained 126 unique records.

The model was built using the MaxEnt algorithm (Phillips et al. 2006). The background was represented by a regular 0.5° grid of 2986 points covering all known distribution of *Aethomys chrysophilus* (our data, Galster et al. 2007, IUCN Red List v. 2015-4) and adjacent areas with similar habitats (Olson et al. 2001). As predictors, we used 19 bioclimatic variables obtained from the WorldClim database (Hijmans et al. 2005). The MaxEnt predictions were expressed as relative occurrence rates (RORs; also called the raw output) divided by uniform prior expectation at each background point $\left(\frac{1}{2986}\right)$. RORs were predicted for all background sites as well as for the corresponding sites in layers containing climate reconstructions for the last glacial maximum (LGM; ~21,000 years BP, Braconnot et al. 2007) and the last inter-glacial (LIG; ~120,000–140,000 years BP, Otto-Bliesner et al. 2006). To evaluate the importance of each predictor we randomized them spatially and calculated the Spearman correlation (r_s) of each predictor before and after randomization. The importance was then quantified as $1–r_s$.

Model selection using the corrected Akaike information criterion (AICc; Warren and Seifert 2011) was employed to choose predictor transformations (so called features) and values of LASSO regularization coefficient which causes some predictors to have zero regression coefficients and thus effectively removes them from the model (see Merow et al. 2013 for full explanation). Both models in the AICc-based confidence set included linear and quadratic features and thus we used them in combination with a weighted mean of the regularization coefficient (=1.31). We did not attempt to reduce the set of bioclimatic variables, but we relied on the LASSO regularization to select their most appropriate subset.

Species distribution modeling was used as implemented in MaxEnt v.3.3.3.k (Phillips et al. 2006) and interfaced to the R computing environment (R Core Team 2016) by using packages "dismo" (Hijmans et al. 2016) and "ENMeval" (Muscarella et al. 2014). The R script is available as the Supplemental File 2 (DOI: 10.6084/m9.figshare.4960406). The results were visualized using the R packages "maptools" (Bivand and Lewin-Koh 2016) and "raster" (Hijmans 2012).

Skull form variation

Morphological data were collected from digital images in the form of landmark configurations covering the skull from its dorsal and ventral side (Figure 2) and described by the set of Cartesian coordinates. The landmark position were digitized in tpsDig 2.18 (Rohlf 2015). The size of each configuration was quantified as a logarithm of its centroid size and its shape was characterized by Procrustes shape coordinates produced by generalized Procrustes analysis (Mitteroecker et al. 2004). The dorsal and ventral skull form (=size and shape) matrices were combined into a single data set and jointly subjected to all analyses.

We tested separately for differences among major phylogeographic lineages and between *Aethomys ineptus* and the rest of the complex by means of partial least squares (PLS) discrimination analyses with skull form as the predictor and binary coded classification as the response.

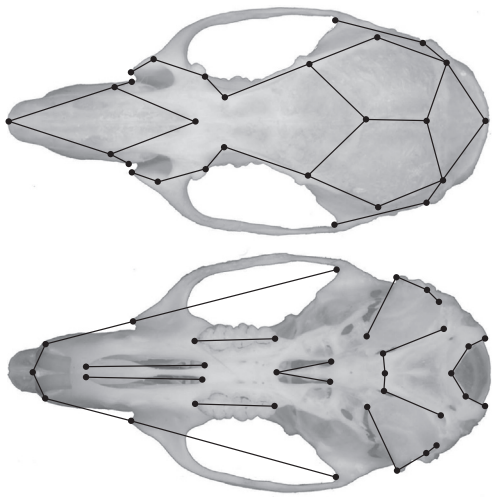


Figure 2: Position of anatomic landmarks on the dorsal and ventral side of the skull.
Landmarks were placed on digital images taken by one of us (OM) in a standardized manner and configurations were then rescaled to units of millimeters. The lines show links between landmarks as displayed when reporting size and shape differences.

More specifically, we used the canonical powered partial least squares (CPPLS) method of Indahl et al. (2009) as implemented in the R package "pls" (Mevik et al. 2015) with the number of components selected by 10-fold cross-validation. The cross-validation was repeated 100 times and we retained the minimum number of components whose median prediction error (PRESS) was within the 50% highest density interval associated with the minimum median PRESS observed. If at least one component was supported, the differentiation was considered significant and analysed in more detail. In total we analysed 44 adult skulls classified according to their *CYTB* barcode or, in a few cases, according to their trapping site (see Supplemental File 1). We included material from our collections as well as from four other museums, namely Ditsong Museum of Natural History (Pretoria, South Africa), Royal Museum for Central Africa (Tervuren, Belgium), Naturmuseum Senckenberg (Frankfurt, Germany) and Smithsonian Institution – National Museum of Natural History (Washington, DC, USA).

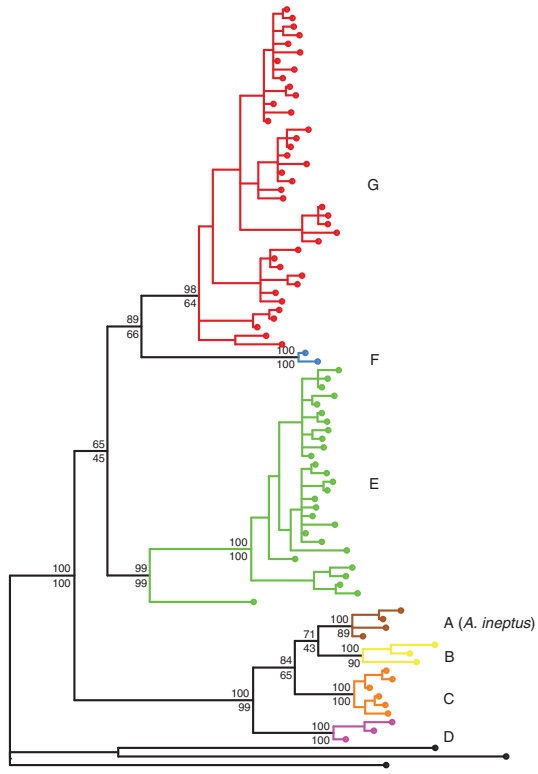


Figure 3: Bayesian tree with nodal values representing: Bayesian posterior probabilities (over the branch) and maximum likelihood bootstrap value (under the branch).
Tip labels are colored as follows: brown for clade A (*A. ineptus*), yellow for clade B, orange for C, violet for D, green for E, blue for F, and red for clade G.

Results

Phylogeny, phylogeographic structure, and divergence dating

Bayesian and maximum likelihood reconstructions of the *CYTB* gene (Figure 3) showed similar topologies with seven highly supported monophyletic clades labeled A to G (posterior probability, PP ≥ 0.98 and bootstrap support, BS ≥ 89 in all cases except for clade G with BS = 64). These units were grouped into three larger groups, A–D (PP = 1.00, BS = 99), F + G (PP = 0.89, BS = 66) and E. Clade E may be a sister clade either to A–D or F + G, where the latter is more probable (PP = 0.65).

All these clades showed largely parapatric distributional ranges (Figure 1) although the actual distributional limits and possible contact zones are unresolved in some cases due to sampling gaps. Clade A–D occupied the southern part of the distributional range, all of them south of the lower Zambezi – Kafue River system. Clade A corresponded to haplotypes from populations designated as *Aethomys ineptus* and its presence is evidenced in the north-east of South Africa where it meets clade B south of the Limpopo River. Clade C was recorded from two sites, one in most western Zimbabwe and the other from the borders of Botswana and Namibia. Clade D was recorded from three sites in southern Zambia, between upper Zambezi and Kafue Rivers. Clade E was documented from central and northern Tanzania and southern-most Kenya, including a distinct haplotype recorded in Namanga Hills (2.53°S, 36.79°E). Clade G occupied a vast area north of the Zambezi River both west and east of Lake Malawi. This clade was not recorded south of the Zambezi River, but this may be due to sampling gaps in Zimbabwe and central Mozambique where only clade F was recorded from a single site in the Chimanimani mountains. In the northern part of the distribution, clades G and E are in contact along the Eastern Arc Mountains and Rukwa Rift in Tanzania. The haplotype network (Figure 4) showed a similar pattern with three very divergent (E–G) and four moderately divergent haplogroups (A–D).

The average phylogenetic distances were 4.01–14.27% between clades and 0.62–2.95% within clades and this gap in distribution of pairwise differences made their

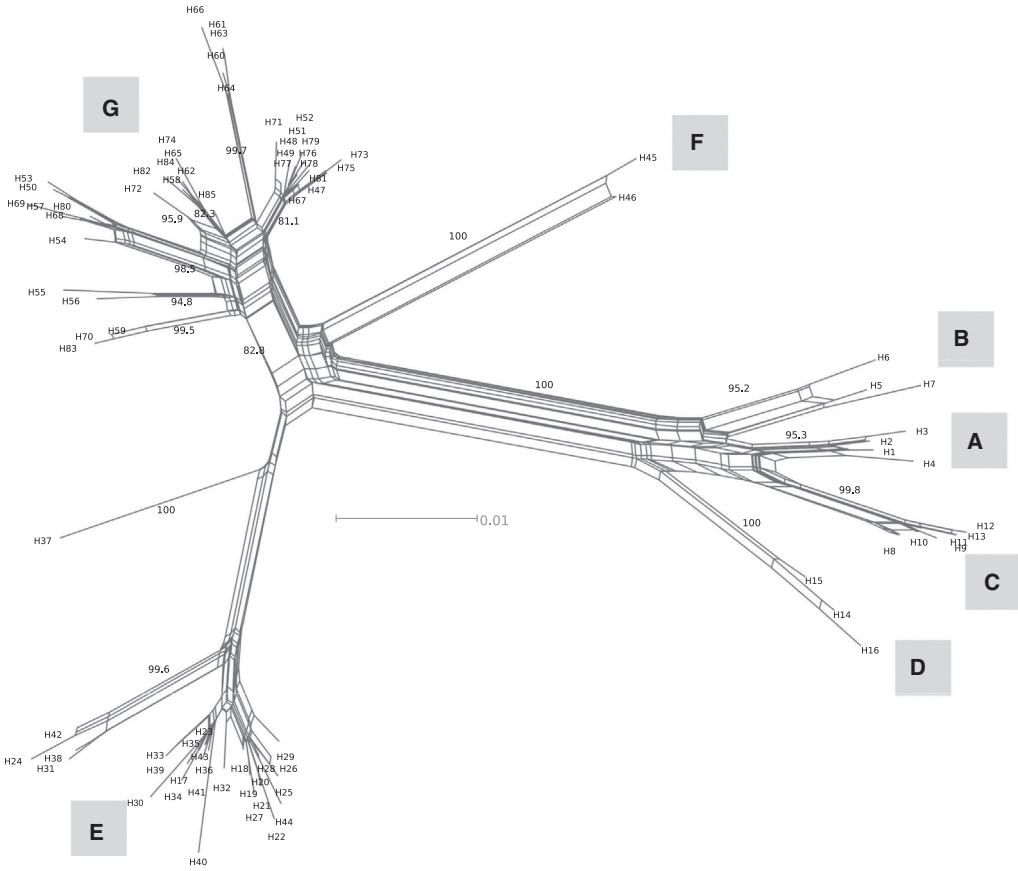


Figure 4: Neighbor-net haplotype network of *CYTB* sequences of the *A. chrysophilus* complex.
Taxon labels represent haplotype definition and letters corresponding clades. K2P distances were used and the network was drawn using the equal angle method. Bootstrap values (1000 replicates) are shown on selected branches.

delimitation possible. Only clades E and G contained more than 10 haplotypes (34 and 70, respectively). Negative values of Tajima's D (–1.84 and –1.10, respectively) indicated non-neutral evolution of DNA variation, whereas negative values of Fu's F_s (–12.12 and –9.88, respectively) suggested genetic hitchhiking or population expansion as the likely causes of non-neutrality for both clades. Population expansion was also supported by bell-shaped mismatch distributions (Figure 5) and Harpending raggedness index (0.068 and 0.0067, respectively). All departures of the statistics from zero expectations (and hence null hypothesis of neutrality) were significant at 0.05 level except for Tajima's D = – 1.10 in clade G (p = 0.12).

The root age of the *Aethomys* phylogeny was estimated at 6.8 Mya with the 95% highest posterior density (HPD) interval as 6–8.6 Mya (Figure 6). Time to the most recent common ancestor for the *Aethomys chrysophilus* complex was estimated at 1.3 (0.5–2.4) Mya and the origin of the *Aethomys ineptus* lineage at 0.7 (0.1–1.4) Mya. The position of *A. ineptus* inside the *A. chrysophilus* complex was supported with a posterior probability value of 0.94.

Species distribution modeling

Our presence records were all included in areas with predicted RORs exceeding uniform prior expectations (Figure 7A) which suggested reasonably good fit of the model. Prediction for the LGM climate produced a distribution map roughly similar to that of current distribution map, albeit more fragmented, especially along the Zambezi River (Figure 7B). On the contrary, prediction for

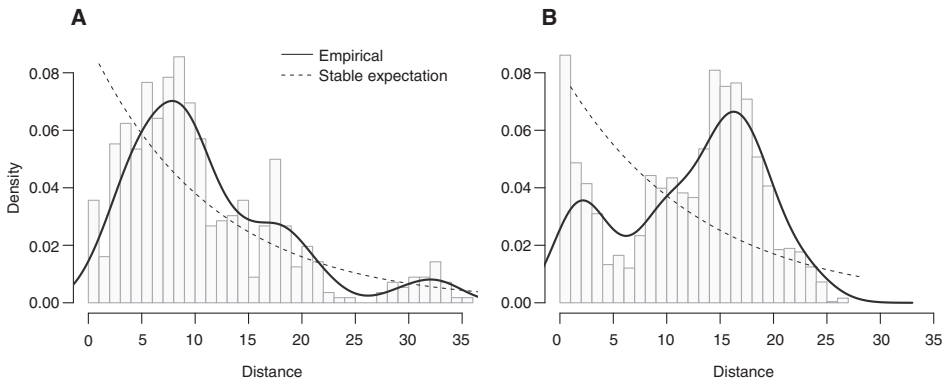


Figure 5: Distribution of the number of pairwise differences among haplotypes in two clades: (A) clade E and (B) clade G of *A. chrysophilus*. Bars and full line represent the observed distribution and the dotted line represents the expected distribution under the model of stable population fitted to the data.

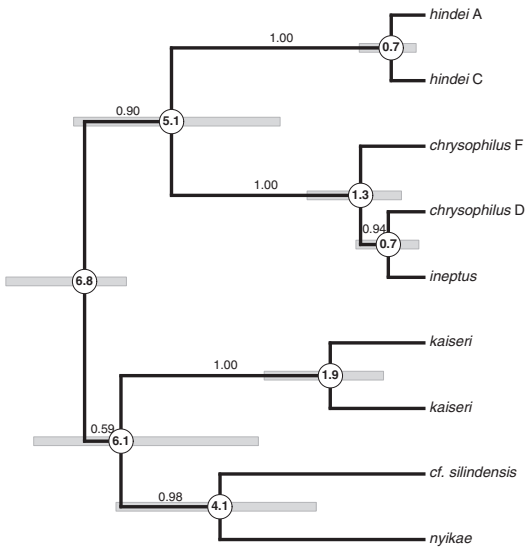


Figure 6: Time calibrated phylogeny of *Aethomys* based on two nuclear genes, *IRBP* and *RAG1*.
Node ages are in Mya, 95% HPD intervals are shown as light gray rectangles. Posterior probabilities of particular clades are shown above branches supporting them. (Designations "*hindei* A" and "*hindei* C" are based on our unpublished phylogeographic data.)

the LIG (Figure 7C) revealed a decrease in suitable conditions across the whole distributional area. In terms of prediction precision loss after randomization the most important variables were: "mean diurnal range of temperatures" (0.68), "mean temperature of driest quarter" (0.65) and "minimum temperature of coldest month" (0.57). According to comparison of predictor distributions with their ROR-weighted versions (Figure 8) *Aethomys chrysophilus* avoids areas with high diurnal range of temperatures, especially where combined with high temperature in the cold and dry season. According to WorldClim data (not shown), diurnal range of temperatures is getting too large (>15°C) in the Kalahari Desert, and the coldest part of the year appears too hot in arid regions of Kenya (min. temperature >12°C).

Skull form differentiation

Skull form differentiation among phylogeographic lineages was examined at a coarse scale by comparing only the three main phylogenetic groups: northern (E), central (F + G) and southern (A to D) with sample sizes 15, 19 and 10, respectively. Cross-validation suggested a single component to be retained, which accounted for about 4% of shape variation, but virtually no size variation in the data set. The cross-validated classification success was low, however, only 67% for the northern lineage, 63% for the central lineage and 0% for the southern one. PLS scores from the final model are shown in Figure 9, suggesting the northern lineage as the most distinct one. Superimposition of shapes predicted for the extreme PLS scores shows the northern lineage as having slightly more robust rostrum and occipital condyles placed more closely to each other (Figure 10). When *Aethomys ineptus* (six skulls) was contrasted against the rest of the complex, no component was supported as significantly improving on the classification success and skulls form of *A. ineptus* showed no differentiation in our analysis.

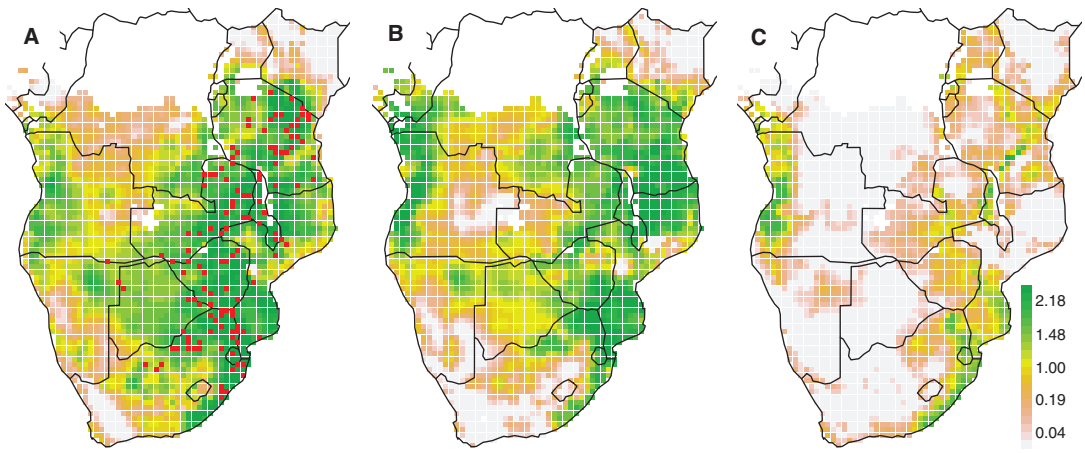


Figure 7: Grid maps of *A. chrysophilus* distribution inferred using maximum entropy model during for (A) current climatic conditions, (B) LGM (~21,000 years BP) and (C) LIG (~120,000–140,000 years BP).
Red squares represent sampled coordinates. Dark green color represents preferred conditions, i.e. high RORs relative to prior expectation.

Discussion and conclusion

Distribution of *Aethomys chrysophilus* complex

Our study confirms that *Aethomys chrysophilus* complex is widely distributed and its habitat requirements are centred mainly to various seasonal savanna-woodland habitats (Skinner and Chimimba 2005, Happold 2013). When suitable habitats are available, these rodents can occur in a wide range of altitudes, from 10 m a.s.l. (Mozambique, Xai-Xai, 21.12°S, 33.74°E) to about 2300 m a.s.l. (Tanzania, Mbizi, 7.87°S, 31.67°E). Our distribution modeling supplements this picture with a climate-based quantitative view, which suggests that in spite of its preference for relatively dry habitats, the species avoids extreme conditions reflected by extreme ambient temperatures of deserts or semi-deserts. This result contradicts occurrence of the species in central Kenya, which is indicated by the IUCN Red List but for which we found no evidence in museum collections. Notably, our model predicts extensive occurrence of the *A. chrysophilus* complex along the Atlantic coast in Angola for which we currently have little evidence for in spite of its presence in Namibia and southern Angola (see e.g. Global Biodiversity Information Facility http://www.gbif.org). It would be interesting to examine whether the species occurs in Angola although undocumented in natural history collections and what are its relationships with south-west African endemic *Aethomys bocagei* (Crawford-Cabral 1998), for which no genetic data are available.

Phylogeographic divides

In this study we found a pronounced phylogeographic structure within the *Aethomys chrysophilus* complex where some clades have been separated by well-known biogeographical divides. The distribution of clade E is delimited by Rukwa Rift and the Eastern Arc Mountains where it meets clade G. These geomorphological features have already been shown to play a prominent role in the history of other rodent taxa living in non-forest habitats such as the pygmy mouse (*Mus minutoides*, Bryja et al. 2014), the multimammate mouse (*Mastomys natalensis*, Colangelo et al. 2013, Gryseels et al. 2017), the silvery mole-rat (*Heliophobius argenteocinereus*, Faulkes et al. 2011) and spiny mice (*Acomys spinosissimus* complex, Verheyen et al. 2011). In each of these rodent species there are phylogeographic lineages separated either by the Eastern Arc Mountains, the Rukwa Rift or both. The same distributional limits were also observed in an African ungulate species, the bushbuck (*Tragelaphus scriptus*, Moodley and Bruford 2007).

The contact between clades D and G is located in the area of Kafue Flats, wetland landscape around the Lower Kafue River (Zambia). This area is also transitional for several other species. In bushveld gerbils (*Gerbilliscus leucogaster*) it forms a barrier to gene flow

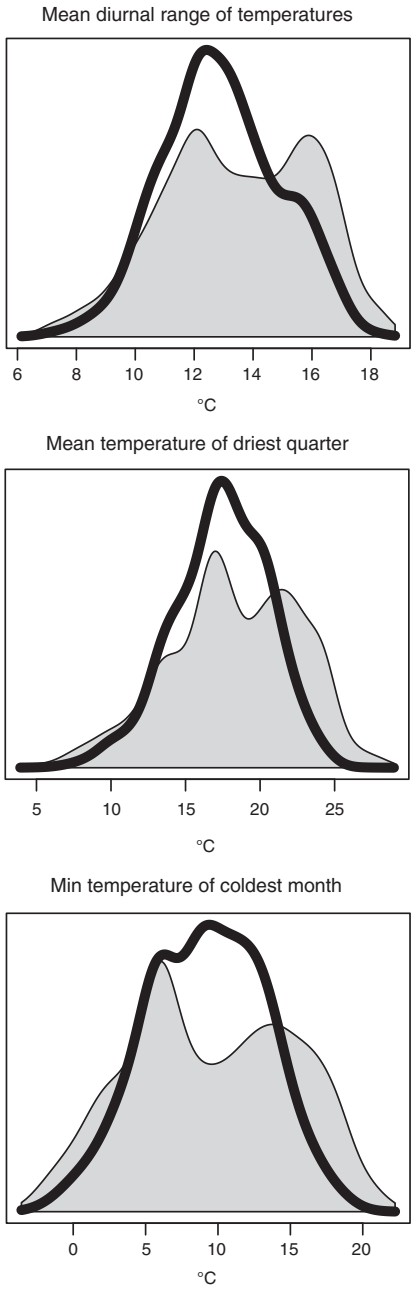


Figure 8: Response plots of three most influential variables in the species distribution model.
Kernel density estimates of background values (in gray) are compared with their ROR-weighted counterparts (in black). In the range of preferred values weighting by the predicted RORs causes an excess of probability density at the expense of less preferred or even avoided intervals.

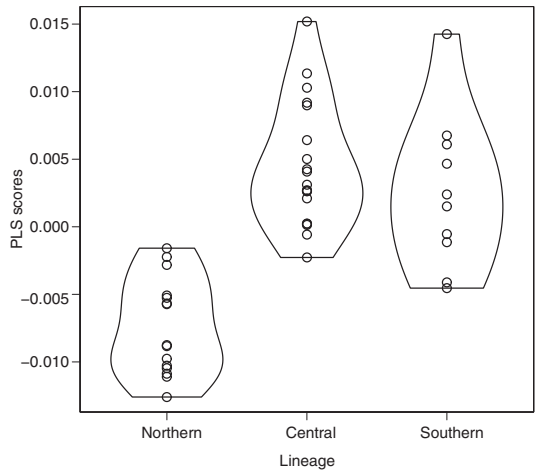


Figure 9: Violin plots showing skull form differentiation among the three major phylogeographic lineages: northern (E), central (F + G) and southern (A to D).
Points correspond to individual PLS scores in the final discrimination analysis, envelopes ("violins") are kernel density estimates of their group-specific distributions produced by vioplot package (Adler 2005).

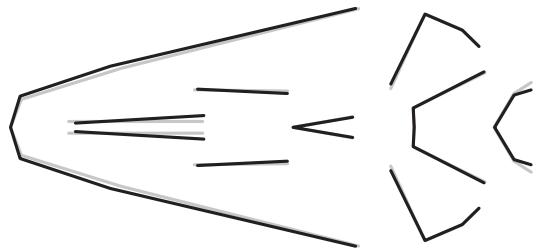


Figure 10: Wireframe plots showing difference in skull shapes (from the ventral view) corresponding to extreme PLS scores.
The shape specific to the northern lineage (in black) is contrasted to the shape specific to the remaining two lineages (in gray).

with observed discordance of nuclear and mitochondrial markers (McDonough et al. 2015) and there is a contact zone between two subspecies of pouched mice (*Saccostomus campestris campestris* and *Saccostomus campestris mashonae*, Mikula et al. 2016). From the perspective of large mammals Kafue Flats are part of a wider biogeographical divide in the distribution of baboons (*Papio*, Zinner et al. 2009) and some ungulates including giraffes (*Giraffa giraffe*, *Giraffa tippelskirchi*, Fennessy et al. 2016) as well as tsessebe antelopes (*Damaliscus*) and wildebeests (*Connochaetes*) as documented by Cotterill (2003).

The association of parapatric contact zones with mountain ridges or big rivers may be due to partial reproductive isolation, because contact zones with selection against hybrids tend to be associated with migration barriers (Barton and Hewitt 1989). In African rodents, this was suggested by evidence for a hybrid zone between two *Saccostomus campestris* subspecies at the Zambezi River (Mikula et al. 2016) and between two lineages of *Mastomys natalensis* along Eastern Arc Mountains (Gryseels et al. 2017). In the *Aethomys chrysophilus* complex it may be relevant especially in the case of the contact zone between clades A and B in the proximity of the Limpopo River where clade A corresponds to putatively distinct *Aethomys ineptus*. The actual contact zone seems to have been shifted southward from the river flow (Linzey et al. 2003, Russo et al. 2006) and worth of further investigation. The other two detected contact zones were between lineages D–G and G–E. The contact zone between lineages G–E along the Eastern Arc Mountains and Rukwa Rift is well delimited by our present sampling which shows one clade replacing the other over tens of kilometers in spite of the lack of obvious migration barriers.

History of the complex

Our molecular dating calibrated by the oldest undisputed *Aethomys* fossil (Lemudong'o, Kenya, 6.1 Mya) and based on two nuclear genes estimated the most recent common ancestor (MRCA) of the genus at 6–8.6 Mya and the MRCA of the *Aethomys chrysophilus* complex, including the putative speciation of *Aethomys ineptus*, was estimated at 0.5–2.4 Mya (Pleistocene age). More precisely, the MRCA of the complex is probably (PP = 0.91) younger than 2.0 Mya when tropical climate switched to its modern mode with hydrological cycles controlled by monsoon dynamics (Ravelo et al. 2004). The changes of the African climate and vegetation since that time were driven by interference of high-latitude and low-latitude forcing. Extensive high-latitude glaciations were associated with low sea surface temperatures and low levels of atmospheric CO_2. This forced the climate to be cold and dry with grasslands expanding at the expense of forest (deMenocal 1995, Schefuss et al. 2003). This resulted in a complex and regionally asynchronous series of habitat changes (Blome et al. 2012, Shanahan et al. 2015, Johnson et al. 2016) with the potential to cause local extinctions and population divergence.

The observed phylogeographic structure within the *Aethomys chrysophilus* complex is attributable to the cumulative effect of this extinction-colonization process. First, both well-sampled clades (E and G) bear signatures of recent population expansion that is either ongoing or young enough to be still apparent in sequence variation. Second, our species distribution model (Figure 7) suggests a range-wide retreat and fragmentation of populations in the LIG (~120,000–140,000 years BP). Putative refugia could have been situated north of the Zambezi River in Victoria basin, west of Lake Malawi and along the Eastern Arc Mountains. Pollen records from Lake Malawi indicates open miombo woodland alternating with grassland during drier and wetter phases of the LIG (Beuning et al. 2011) which should support a permanent population of *Aethomys chrysophilus* in this area. At the same time, forests could separate these putative refugia as they are documented e.g. from the Rukwa rift even during the LGM (Vincens et al. 2005). More extensive inter-glacial distributions are predicted around the Kafue and Luangwa Rivers, south of the Lower Zambezi River and around the Limpopo River including the south.

Skull shape differentiation

We also found some evidence for skull shape differentiation between the three major phylogeographic lineages. Individuals from the northern lineage (E) inhabiting arid habitats in northwestern Tanzania and southern Kenya had more robust rostrum and differently shaped occipital regions. In terms of cross-validated classification success the differentiation was moderate (67%) but significant (PRESS of the null model was outside its 50% highest density interval). We suggest, therefore, the contrast between lineage E and the rest as the most likely case of local adaptation within the complex. This should be investigated further by using multiple functional traits.

Taxonomic status of *Aethomys ineptus*

Aethomys ineptus is now considered a valid species (Musser and Carleton 2005, Chimimba and Linzey 2008) based on its 2n = 44 karyotype, aberrant sperm morphology and mitochondrial monophyly. In the rest of the *Aethomys chrysophilus* complex as well as in other studied species of *Aethomys* the diploid chromosome number is 50 (Matthey 1958) and spermatozoa resemble those of other murines (Breed 2004). From a systematic and taxonomic point of view *A. ineptus* may thus represent an interesting case of a species that is monophyletic, defined by clearly derived traits, but phylogenetically embedded in its parental species (a "budding speciation" scenario of Funk and Omland 2003).

Spermatozoa of *Aethomys ineptus* differ in many respects including the spatulate-shaped sperm head with no apical hook, modified ultrastructure and shorter tails (Gordon and Watson 1986, Visser and Robinson 1987, Breed et al. 1988). Given the importance of these characteristics in fertilization and postcopulatory selection (Gómez Montoto et al. 2011, Varea-Sánchez et al. 2014) these features may act as a reproductive barrier. The karyotype of *A. ineptus* differs by three centric fusions (Visser and Robinson 1986) that are, by themselves, less likely to cause reproductive incompatibility (Maputla et al. 2011, Dobigny et al. 2015, Medarde et al. 2015), but minor structural changes were also observed. No putative hybrid with intermediate diploid number or any conflicting trait combination has ever been observed (Linzey et al. 2003) which suggests reproductive isolation. On the other hand, distributional patterns of mtDNA and karyotypes are not fully concordant which suggest a degree of genomic admixture. More specifically, both karyotypes were found within the range of our clade A (Visser and Robinson 1986) as well as far beyond it (namely in Zimbabwe, Gordon and Rautenbach 1980). Finally, there may be a mechanistic link between presence of centric fusions and sperm morphology (Medarde et al. 2013) which would make the coincidence of these traits in *A. ineptus* less surprising but possibly more important.

Morphological traits suggested earlier as diagnostic were not found to be reliable in later studies. *Aethomys ineptus* has no specific shape of the baculum (Visser and Robinson 1987) and the craniometric delimitation of the species (Chimimba et al. 1999) is in conflict with other evidence (Linzey et al. 2003, this study). Our present analysis also suggests no differentiation in skull size and shape between *A. ineptus* and the rest of the complex. Future studies employing multiple nuclear markers may help to reveal the level of reproductive isolation in the contact zone of both taxa in southern Africa.

Acknowledgments: This study was supported by two projects of the Czech Science Foundation, nos. P506/10/0983 and 15-20229S. We are very grateful to Janice Britton-Davidian and Molly McDonough for providing access to their valuable material. We are indebted to many local authorities for providing necessary permits to carry out our research, especially Zambia Wildlife Authority (ZAWA), National Directorate for Protected Areas (DINAC – Mozambique), the Kenyan Forest Service and the Kenyan Wildlife Service, and the Sokoine University of Agriculture in Morogoro, Tanzania. T. Aghová and A. Bryjová helped with genotyping. We express our thanks to our colleagues J. Šklíba, M. Lövy, E. Hrouzková, O. Kott, S. Gryseels, J. Vrbová Komárková, J. Krásová, G. Phamphi, C. S. Sabuni, A. Katakweba, A. Massawe, as well as protected area officers, rangers and countless local assistants, for the much-needed support during field expeditions.

References

Adler, D. 2005. vioplot: Violin plot. R package version 0.2. http://wsopuppenkiste.wiso.uni-goettingen.de/~dadler.

Aghová, T., R. Šumbera, L. Piálek, O. Mikula, M.M. McDonough, L.A. Lavrenchenko, Y. Meheretu, J.S. Mbau and J. Bryja. 2017. Multilocus phylogeny of east African gerbils (Rodentia, Gerbilliscus) illuminates the history of the Somali-Masai savanna. J. Biogeogr. in press, doi:10.1111/jbi.13017.

Baker, R.J., M.B. Qumsiyeh and I.L. Rautenbach. 1988. Evidence for eight tandem and five centric fusions in the evolution of the karyotype of *Aethomys namaquensis* A. Smith (Rodentia: Muridae). Genetica 76: 161–169.

Barton, N.H. and G.M. Hewitt. 1989. Adaptation, speciation and hybrid zones. Nature 341: 497–503.

Berger, S.A., D. Krompass and A. Stamatakis. 2011. Performance, accuracy, and web server for evolutionary placement of short sequence reads under maximum likelihood. Syst. Biol. 60: 291–302.

Beuning, K.R.M., K.A. Zimmerman, S.J. Ivory and A.S. Cohen. 2011. Vegetation response to glacial-interglacial climate variability near Lake Malawi in the southern African tropics. Palaeogeogr. Palaeoclimatol. Palaeoecol. 303: 81–92.

Bivand, R. and N. Lewin-Koh. 2016. maptools: tools for Reading and Handling Spatial Objects. https://CRAN.R-project.org/package=maptools.

Blome, M.W., A.S. Cohen, C.A. Tryon, A.S. Brooks and J. Russell. 2012. The environmental context for the origins of modern human diversity: a synthesis of regional variability in African climate 150,000–30,000 years ago. J. Hum. Evol. 62: 563–592.

Bohoussou, K.H., R. Cornette, B. Akpatou, M. Colyn, J.C. Kerbis Peterhans, J. Kennis, R. Šumbera, E. Verheyen, E. N'Goran, P. Katuala and V. Nicolas. 2015. The phylogeography of the rodent genus *Malacomys* suggests multiple Afrotropical Pleistocene lowland forest refugia. J. Biogeogr. 42: 2049–2061.

Bouckaert, R.R., J. Heled, D. Kühnert, T. Vaughan, C.H. Wu, D. Xie, M.A. Suchard, A. Rambaut and A.J. Drummond. 2014. BEAST 2: A Software Platform for Bayesian Evolutionary Analysis. PLoS Comput. Biol. 10: e1003537.

Braconnot, P., B. Otto-Bliesner, S. Harrison, S. Joussaume, J.-Y. Peterchmitt, A. Abe-Ouchi, M. Crucifix, E. Driesschaert, T. Fichefet, C.D. Hewitt, M. Kageyama, A. Kitoh, A. Laîné, M.-F. Loutre, O. Marti, U. Merkel, G. Ramstein, P. Valdes, S.L. Weber, Y. Yu and Y. Zhao. 2007. Results of PMIP2 coupled simulations of the Mid-Holocene and Last Glacial Maximum – Part 1: experiments and large-scale features. Clim. Past. 3: 261–277.

Breed, W.G. 2004. The spermatozoon of eurasian murine rodents: Its morphological diversity and evolution. J. Morphol. 261: 52–69.

Breed, W.G., G.A. Cox, C.M. Leigh and P. Hawkins. 1988. Sperm head structure of a murid rodent from southern Africa: the red veld rat *Aethomys chrysophilus*. Gamete Res. 19: 191–202.

Bryja, J., L. Granjon, G. Dobigny, H. Patzenhauerová, A. Konečný, J.M. Duplantier, P. Gauthier, M. Colyn, L. Durnez, A. Lalis and V. Nicolas. 2010. Plio-Pleistocene history of West African Sudanian savanna and the phylogeography of the *Praomys daltoni* complex (Rodentia): the environment/geography/genetic interplay. Mol. Ecol. 19: 4783–4799.

Bryja, J., O. Mikula, R. Šumbera, Y. Meheretu, T. Aghová, L.A. Lavrenchenko, V. Mazoch, N. Oguge, J.S. Mbau, K. Welegerima, N. Amundala, M. Colyn, H. Leirs and E. Verheyen. 2014. Pan-African phylogeny of *Mus* (subgenus *Nannomys*) reveals one of the most successful mammal radiations in Africa. BMC Evol. Biol. 14: 256.

Castiglia, R., M. Corti, P. Colangelo, F. Annesi, E. Capanna, W. Verheyen, A.M. Sichilima and R. Makundi. 2003. Chromosomal and molecular characterization of *Aethomys kaiseri* from Zambia and *Aethomys chrysophilus* from Tanzania (Rodentia, Muridae). Hereditas. 139: 81–89.

Chimimba, C.T. 1998. A taxonomic synthesis of southern African *Aethomys* (Rodentia: Muridae) with a key to species. Mammalia 62: 427–438.

Chimimba, C.T. and A. V. Linzey. 2008. *Aethomys ineptus* (Rodentia: Muridae). Mamm. Species. 809: 1–7.

Chimimba, C.T., N.J. Dippenaar and T.J. Robinson. 1999. Morphometric and morphological delineation of southern African species of *Aethomys* (Rodentia: Muridae). Biol. J. Linn. Soc. 67: 501–527.

Colangelo, P., E. Verheyen, H. Leirs, C. Tatard, C. Denys, G. Dobigny, J.M. Duplantier, C. Brouat, L. Granjon and E. Lecompte. 2013. A mitochondrial phylogeographic scenario for the most widespread African rodent, *Mastomys natalensis*. Biol. J. Linn. Soc. 108: 901–916.

Cotterill, F.P.D. 2003. A biogeographic review of tsessebe antelopes, *Damaliscus lunatus* (Bovidae: Alcelaphini), in south-central Africa. Durban Museum Novit. 28: 45–55.

Crawford-Cabral, J. 1998. The Angolan rodents of the superfamily Muroidea: an account of their distribution. Instituto de Investigação Científica Tropical, Estudos, Ensaios e Documentos, Vol. 161, pp. 1–222.

Darriba, D., G.L. Taboada, R. Doallo and D. Posada. 2012. jModelTest 2: more models, new heuristics and parallel computing. Nat. Methods. 9: 772.

Deino, A.L. and S.H. Ambrose. 2007. 40Ar/39Ar dating of the Lemudong'o late Miocene fossil assemblages, southern Kenya rift. Kirtlandia. 56: 65–71.

deMenocal, P.B. 1995. Plio-Pleistocene African climate. Science 270: 53–59.

Demos, T.C., J.C. Kerbis Peterhans, B. Agwanda and M.J. Hickerson. 2014. Uncovering cryptic diversity and refugial persistence among small mammal lineages across the Eastern Afromontane biodiversity hotspot. Mol. Phylogenet. Evol. 71: 41–54.

Dobigny, G., C. Tatard, P. Gauthier, K. Ba, J.-M. Duplantier, L. Granjon and G.J. Kergoat. 2013. Mitochondrial and nuclear genes-based phylogeography of *Arvicanthis niloticus* (Murinae) and sub-Saharan open habitats pleistocene history. PLoS One 8: e77815.

Dobigny, G., J. Britton-Davidian and T.J. Robinson. 2015. Chromosomal polymorphism in mammals: an evolutionary perspective. Biol. Rev. 92: 1–21.

Drummond, A.J., S.Y.W. Ho, M.J. Phillips and A. Rambaut. 2006. Relaxed Phylogenetics and Dating with Confidence. PLoS Biol. 4: e88.

Ducroz, J.F., V. Volobouev and L. Granjon. 2001. An assessment of the systematics of Arvicanthine rodents using mitochondrial DNA sequences: evolutionary and biogeographical implications. J. Mamm. Evol. 8: 173–206.

Fadda, C., R. Castiglia, P. Colangelo, M. Corti, R. Machang'u, A. Scanzani, P. Tesha, W. Verheyen and E. Capanna. 2001. The rodent fauna of Tanzania: a cytotaxonomic report from the Maasai Steppe (1999). Rend. Fis. Acc. Lincei 9: 29–49.

Faulkes, C.G., N.C. Bennett, F.P.D. Cotterill, W. Stanley, G.F. Mgode and E. Verheyen. 2011. Phylogeography and cryptic diversity of the solitary-dwelling silvery mole-rat, genus *Heliophobius* (family: Bathyergidae). J. Zool. 285: 324–338.

Fennessy, J., T. Bidon, F. Reuss, V. Kumar, P. Elkan, M.A. Nilsson, M. Vamberger, U. Fritz and A. Janke. 2016. Multi-locus analyses reveal four giraffe species instead of one. Curr. Biol. 26: 1–7.

Fu, Y.X. 1997. Statistical tests of neutrality of mutations against population growth, hitchhiking and background selection. Genetics. 147: 915–925.

Funk, D.J. and K.E. Omland. 2003. Species-level paraphyly and polyphyly: frequency, causes, and consequences, with insights from animal mitochondrial DNA. Annu. Rev. Ecol. Evol. Syst. 34: 397–423.

Galan, M., M. Pagès and J.F. Cosson. 2012. Next-generation sequencing for rodent barcoding: species identification from fresh, degraded and environmental samples. PLoS One 7: e48374.

Galster, S., N.D. Burgess, J. Fjeldså, L.A. Hansen and C. Rahbek. 2007. One degree resolution databases of the distribution of 1085 species of mammals in Sub-Saharan Africa. On-line data source-Version 1.00. Zoological Museum, University of Copenhagen, Denmark. Available via http://www.daim.snm.ku.dk/african-vertebrates-searc.

Geraads, D., Z. Alemseged and H. Bellon. 2002. The late Miocene mammalian fauna of Chorora, Awash basin, Ethiopia: systematics, biochronology and 40K–40Ar ages of the associated volcanics. Tert. Res. 21: 113–122.

Gómez Montoto, L., C. Magaña, M. Tourmente, J. Martín-Coello, C. Crespo, J.J. Luque-Larena, M. Gomendio and E.R.S. Roldan. 2011. Sperm competition, sperm numbers and sperm quality in muroid rodents. PLoS One 6: e18173.

Gordon, D.H. and I.L. Rautenbach. 1980. Species complexes in medically important rodents: chromosome studies of *Aethomys*, *Tatera* and *Saccostomus* (Rodentia: Muridae, Cricetidae). S. Afr. J. Sci. 76: 559–561.

Gordon, D.H. and C.R.B. Watson. 1986. Identification of cryptic species of rodents (*Mastomys*, *Aethomys*, *Saccostomus*) in the Kruger National Park. South African J. Zool. 21: 95–99.

Gryseels, S., S.J.E. Baird, B. Borremans, R. Makundi, H. Leirs and J. Goüy de Bellocq. 2017. When viruses don't go viral: the importance of host phylogeographic structure in the spatial spread of arenaviruses. PLoS Pathogens 13: e1006073.

Happold, D.C.D. (ed.). 2013. Volume III. Rodents, Hares and Rabbits. Bloomsbury Publishing, London. p. 789.

Haus, T., E. Akom, B. Agwanda, M. Hofreiter, C. Roos and D. Zinner. 2013. Mitochondrial diversity and distribution of African Green Monkeys (*Chlorocebus* Gray, 1870). Am. J. Primatol. 75: 350–360.

Heled, J. and A.J. Drummond. 2012. Calibrated tree priors for relaxed phylogenetics and divergence time estimation. Syst. Biol. 61: 138–149.

Hijmans, R.J. 2012. Introduction to the 'raster' package (version 1.9–63). https://CRAN.R-project.org/package=raster.
Hijmans, R.J., S.E. Cameron, J.L. Parra, P.G. Jones and A. Jarvis. 2005. Very high resolution interpolated climate surfaces for global land areas. Int. J. Climatol. 25: 1965–1978.
Hijmans, R.J., S. Phillips, J. Leathwick and J. Elith. 2016. Dismo: species distribution modeling. https://CRAN.R-project.org/package=dismo.
Huson, D.H. and D. Bryant. 2006. Application of phylogenetic networks in evolutionary studies. Mol. Biol. Evol. 23: 254–267.
Indahl, U.G., K.H. Liland and T. Næs. 2009. Canonical partial least squares-a unified PLS approach to classification and regression problems. J. Chemom. 23: 495–504.
Irwin, D.M., T.D. Kocher and A.C. Wilson. 1991. Evolution of cytochrome-b gene in mammals. J. Mol. Evol. 32: 128–144.
Johnson, T.C., J.P. Werne, E.T. Brown, A. Abbott, M. Berke, J. Halbur, S. Contreras – Quintana, S. Grossheusch, S. Schouten, J. Sinninghe Damsté, R. Lyons, C.A. Scholz, A. Cohen and J. King. 2016. A progressively wetter climate in Southern East Africa over the past million years. Quat. Int. 404: 174–175.
Kimura, M. 1980. A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. J. Mol. Evol. 16: 111–120.
Lehner, B. and P. Döll. 2004. Development and validation of a global database of lakes, reservoirs and wetlands. J. Hydrol. 296: 1–22.
Librado, P. and J. Rozas. 2009. DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. Bioinformatics. 25: 1451–1452.
Linzey, A.V. and C.T. Chimimba. 2008. *Aethomys chrysophilus* (Rodentia: Muridae). Mamm. Species. 808: 1–10.
Linzey, A.V., M.H. Kesner, C.T. Chimimba and C. Newbery. 2003. Distribution of veld rat sibling species *Aethomys chrysophilus* and *Aethomys ineptus* (Rodentia: Muridae) in southern Africa. African Zool. 38: 169–174.
Lorenzen, E.D., R. Heller and H.R. Siegismund. 2012. Comparative phylogeography of African savannah ungulates. Mol. Ecol. 21: 3656–3670.
Manthi, F.K. 2007. A preliminary review of the rodent fauna from Lemudong'o, southwestern Kenya, and its implication to the late Miocene paleoenvironments. Kirtlandia. 56: 92–105.
Maputla, N.W., E.R. Dempster, J. Raman and J.W.H. Ferguson. 2011. Strong hybrid viability between two widely divergent chromosomal forms of the pouched mouse. J. Zool. 285: 180–187.
Matthey, R. 1958. Les chromosomes et la position systématique de quelques Murinae africains (Mammalia Rodentia). Acta Trop. 15: 7–117.
McDonough, M.M., R. Šumbera, V. Mazoch, A.W. Ferguson, C.D. Phillips and J. Bryja. 2015. Multilocus phylogeography of a widespread savanna-woodland adapted rodent reveals the influence of Pleistocene geomorphology and climate change in Africa's Zambezi region. Mol. Ecol. 24: 5248–5266.
Medarde, N., J. Martínez-Vargas, A. Sánchez-Chardi, M.J. López-Fuster and J. Ventura. 2013. Effect of Robertsonian translocations on sperm head form in the house mouse. Biol. J. Linn. Soc. 110: 878–889.
Medarde, N., V. Merico, M.J. López-Fuster, M. Zuccotti, S. Garagna and J. Ventura. 2015. Impact of the number of Robertsonian chromosomes on germ cell death in wild male house mice. Chromosom. Res. 23: 159–169.
Mein, P., M. Pickford and B. Senut. 2004. Late Miocene micromammals from the Harasib karst deposits, Namibia. Pt. 2b. Cricetomyidae, Dendromuridae and Muridae, with an addendum on the Myocricetodontinae. Commun. Geol. Surv. Namibia. 13: 43–62.
Merow, C., M.J. Smith and J.A. Silander. 2013. A practical guide to MaxEnt for modeling species' distributions: what it does, and why inputs and settings matter. Ecography (Cop.). 36: 1058–1069.
Mevik, B.-H., R. Wehrens and K.H. Liland. 2015. pls: Partial Least Squares and Principal Component Regression. https://CRAN.R-project.org/package=pls.
Mikula, O., R. Šumbera, T. Aghová, J.S. Mbau, A.S. Katakweba, C.A. Sabuni and J. Bryja. 2016. Evolutionary history and species diversity of African pouched mice (Rodentia: Nesomyidae: *Saccostomus*). Zool. Scr. 45: 595–617.
Miller, M.A., W. Pfeiffer and T. Schwartz. 2010. Creating the CIPRES Science Gateway for inference of large phylogenetic trees. In: (Gateway Computing Environments Workshop (GCE), 2010). IEEE, New Orleans, LA, USA, pp. 1–8.
Mitteroecker, P., P. Gunz, M. Bernhard, K. Schaefer and F.L. Bookstein. 2004. Comparison of cranial ontogenetic trajectories among great apes and humans. J. Hum. Evol. 46: 679–698.
Moodley, Y. and M.W. Bruford. 2007. Molecular biogeography: towards an integrated framework for conserving Pan-African biodiversity. PLoS One 2: e454.
Muscarella, R., P.J. Galante, M. Soley-Guardia, R.A. Boria, J.M. Kass, M. Uriarte and R.P. Anderson. 2014. ENMeval: an R package for conducting spatially independent evaluations and estimating optimal model complexity for MAXENT ecological niche models. Methods Ecol. Evol. 5: 1198–1205.
Musser, G.G. and M.D. Carleton. 2005. Superfamily Muroidea. In: (D.E., Wilson and D.M. Reeder, eds.) Mammal species of the world: a taxonomic and geographic reference. 3rd ed. Johns Hopkins University Press, Baltimore, MD. pp. 894–1531.
Nicolas, V., J.F. Mboumba, E. Verheyen, C. Denys, E. Lecompte, A. Olayemi, A.D. Missoup, P. Katuala and M. Colyn. 2008. Phylogeographic structure and regional history of *Lemniscomys striatus* (Rodentia: Muridae) in tropical Africa. J. Biogeogr. 35: 2074–2089.
Nicolas, V., A.D. Missoup, C. Denys, J.C. Kerbis Peterhans, P. Katuala, A. Couloux and M. Colyn. 2011. The roles of rivers and Pleistocene refugia in shaping genetic diversity in *Praomys misonnei* in tropical Africa. J. Biogeogr. 38: 191–207.
Olson, D.M., E. Dinerstein, E.D. Wikramanayake, N.D. Burgess, G.V.N. Powell, E.C. Underwood, J.A. D'amico, I. Itoua, H.E. Strand, J.C. Morrison, C.J. Loucks, T.F. Allnutt, T.H. Ricketts, Y. Kura, J.F. Lamoreux, W.W. Wettengel, P. Hedao, K.R. Kassem, E.D. Wikramanaya, N.D. Burgess, G.V.N. Powell, E.C. Underwood, J.A. D'amico, I. Itoua, H.E. Strand, J.C. Morrison, C.J. Loucks, T.F. Allnutt, T.H. Ricketts, Y. Kura, J.F. Lamoreux, W.W. Wettengel, P. Hedao and K.R. Kassem. 2001. Terrestrial ecoregions of the world: a new map of life on earth. Bioscience. 51: 933.
Otto-Bliesner, B.L., S.J. Marshall, J.T. Overpeck, G.H. Miller and A. Hu. 2006. Simulating Arctic climate warmth and icefield retreat in the last interglaciation. Science 311: 1751–1753.
Paradis, E. 2010. Pegas: an R package for population genetics with an integrated–modular approach. Bioinformatics 26: 419–420.

Paradis, E., J. Claude and K. Strimmer. 2004. APE: analyses of phylogenetics and evolution in {R} language. Bioinformatics 20: 289–290.
Phillips, S.J., R.P. Anderson and R.E. Schapire. 2006. Maximum entropy modeling of species geographic distributions. Ecol. Modell. 190: 231–259.
Phukuntsi, M.A., H. Brettschneider, D.L. Dalton, T. Kearney, J. Badenhorst and A. Kotze. 2016. DNA barcoding for identification of cryptic species in the field and existing museum collections: a case study of *Aethomys* and *Micaelamys* (Rodentia: Muridae). African Zoology 2016: 1–8.
Rambaut, A. and A.J. Drummond. 2013. Tracer 1.6. University of Edinburgh, Edinburgh.
Ravelo, A.C., D.H. Andreasen, M. Lyle, A. Olivarez Lyle and M.W. Wara. 2004. Regional climate shifts caused by gradual global cooling in the Pliocene epoch. Nature 429: 263–267.
R Core Team. 2016. R: a language and environment for statistical computing.
Rogers, A.R. and H. Harpending. 1992. Population growth makes waves in the distribution of pairwise genetic differences. Mol. Biol. Evol. 9: 552–569.
Rohlf, F.J. 2015. tpsDig 2.18 Department of Ecology and Evolution, State University of New York, Stony Brook. Avalibale at: http://life.bio.sunysb.edu/morph.
Ronquist, F., M. Teslenko, P. van der Mark, D.L. Ayres, A. Darling, S. Höhna, B. Larget, L. Liu, M.A. Suchard and J.P. Huelsenbeck. 2012. MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. Syst. Biol. 61: 539–542.
Russo, I.M., C.T. Chimimba and P. Bloomer. 2006. Mitochondrial DNA differentiation between two species of *Aethomys* (Rodentia: Muridae) from southern Africa. J. Mammal. 87: 545–553.
Russo, I.M., C.T. Chimimba and P. Bloomer. 2010. Bioregion heterogeneity correlates with extensive mitochondrial DNA diversity in the Namaqua rock mouse, *Micaelamys namaquensis* (Rodentia: Muridae) from southern Africa – evidence for a species complex. BMC Evol. Biol. 10: 307.
Schefuss, E., S. Schouten, J.H.F. Jansen and J.S.S. Damsté. 2003. African vegetation controlled by tropical sea surface temperatures in the mid-Pleistocene period. Nature 422: 418–421.
Shanahan, T.M., N.P. McKay, K.A. Hughen, J.T. Overpeck, B. Otto-Bliesner, C.W. Heil, J. King, C.A. Scholz and J. Peck. 2015. The time-transgressive termination of the African Humid Period. Nat. Geosci. 8: 140–144.
Skinner, J.D. and C.T. Chimimba. 2005. The mammals of the southern African sub-region (3rd ed.). Cambridge University Press, Cambridge. p. 872.
Stamatakis, A. 2014. RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics 30: 1312–1313.
Stamatakis, A., P. Hoover and J. Rougemont. 2008. A rapid bootstrap algorithm for the RAxML Web servers. Syst. Biol. 57: 758–771.
Stanhope, M.J., J. Czelusniak, J.-S. Si, J. Nickerson and M. Goodman. 1992. A molecular perspective on mammalian evolution from the gene encoding interphotoreceptor retinoid binding protein, with convincing evidence for bat monophyly. Mol. Phylogenet. Evol. 1: 148–160.
Suwa, G., Y. Beyene, H. Nakaya, R.L. Bernor, J.-R. Boisserie, F. Bibi, S.H. Ambrose, K. Sano, S. Katoh and B. Asfaw. 2015. Newly discovered cercopithecid, equid and other mammalian fossils from the Chorora Formation, Ethiopia. Anthropol. Sci. 123: 19–39.
Tajima, F. 1989. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. Genetics 123: 585–595.
Teeling, E.C., M. Scally, D.J. Kao, M.L. Romagnoli, M.S. Springer, and M.J. Stanhope. 2000. Molecular evidence regarding the origin of echolocation and flight in bats. Nature 403: 188–192.
Varea-Sánchez, M., L. Gómez Montoto, M. Tourmente and E.R.S. Roldan. 2014. Postcopulatory sexual selection results in spermatozoa with more uniform head and flagellum sizes in rodents. PLoS One 9: e108148.
Verheyen, W., J. Hulselmans, W. Wendelen, H. Leirs and M. Corti. 2011. Contribution to the systematics and zoogeography of the East-African *Acomys*. Zootaxa. 3059: 1–35.
Vincens, A., G. Buchet, D. Williamson and M. Taieb. 2005. A 23,000 yr pollen record from Lake Rukwa (8°S, SW Tanzania): New data on vegetation dynamics and climate in Central Eastern Africa. Rev. Palaeobot. Palynol. 137: 147–162.
Visser, D.S. and T.J. Robinson. 1986. Cytosystematics of the South African *Aethomys* (Rodentia: Muridae). South African J. Zool. 21: 264–268.
Visser, D.S. and T.J. Robinson. 1987. Systematic implications of spermatozoan and bacular morphology for the South African *Aethomys*. Mammalia 51: 447–454.
Warren, D.L. and S.N. Seifert. 2011. Ecological niche modeling in Maxent: the importance of model complexity and the performance of model selection criteria. Ecol. Appl. 21: 335–342.
Zinner, D., L.F. Groeneveld, C. Keller and C. Roos. 2009. Mitochondrial phylogeography of baboons (*Papio* spp.): indication for introgressive hybridization? BMC Evol. Biol. 9: 83.
Zinner, D., J. Wertheimer, R. Liedigk, L.F. Groeneveld and C. Roos. 2013. Baboon phylogeny as inferred from complete mitochondrial genomes. Am. J. Phys. Anthropol. 150: 133–140.

Supplemental Material: The online version of this article offers supplementary material (https://doi.org/10.6084/m9.figshare.4516745 and https://doi.org/10.6084/m9.figshare.4960406).

Chapter II

Multilocus phylogeography of a widespread savanna-woodland-adapted rodent reveals the influence of Pleistocene geomorphology and climate change in Africa's Zambezi region.

McDonough M. M., Šumbera R., **Mazoch V.**, Ferguson A. W., Phillips C. D., & Bryja J. (2015). *Molecular Ecology*, 24(20), 5248-5266.

MOLECULAR ECOLOGY

Molecular Ecology (2015) **24**, 5248–5266 doi: 10.1111/mec.13374

Multilocus phylogeography of a widespread savanna–woodland-adapted rodent reveals the influence of Pleistocene geomorphology and climate change in Africa's Zambezi region

MOLLY M. MCDONOUGH,*† RADIM ŠUMBERA,‡ VLADIMÍR MAZOCH,‡ ADAM W. FERGUSON,* CALEB D. PHILLIPS*§ and JOSEF BRYJA¶**

*Department of Biological Sciences, Texas Tech University, Lubbock, TX, USA, †National Museum of Natural History, Smithsonian Institution, Washington, DC, USA, ‡Department of Zoology, Faculty of Science, University of South Bohemia, České Budějovice, Czech Republic, §Research and Testing Laboratory, Lubbock, TX, USA, ¶Institute of Vertebrate Biology, Academy of Sciences of the Czech Republic, Brno, Czech Republic, **Department of Botany and Zoology, Faculty of Science, Masaryk University, Brno, Czech Republic

Abstract

Understanding historical influences of climate and physiographic barriers in shaping patterns of biodiversity remains limited for many regions of the world. For mammals of continental Africa, phylogeographic studies, particularly for West African lineages, implicate both geographic barriers and climate oscillations in shaping small mammal diversity. In contrast, studies for southern African species have revealed conflicting phylogenetic patterns for how mammalian lineages respond to both climate change and geologic events such as river formation, especially during the Pleistocene. However, these studies were often biased by limited geographic sampling or exclusively focused on large-bodied taxa. We exploited the broad southern African distribution of a savanna–woodland-adapted African rodent, *Gerbilliscus leucogaster* (bushveld gerbil) and generated mitochondrial, autosomal and sex chromosome data to quantify regional signatures of climatic and vicariant biogeographic phenomena. Results indicate the most recent common ancestor for all *G. leucogaster* lineages occurred during the early Pleistocene. We documented six divergent mitochondrial lineages that diverged ~0.270–0.100 mya, each of which was geographically isolated during periods characterized by alterations to the course of the Zambezi River and its tributaries as well as regional 'megadroughts'. Results demonstrate the presence of a widespread lineage exhibiting demographic expansion ~0.065–0.035 mya, a time that coincides with savanna–woodland expansion across southern Africa. A multilocus autosomal perspective revealed the influence of the Kafue River as a current barrier to gene flow and regions of secondary contact among divergent mitochondrial lineages. Our results demonstrate the importance of both climatic fluctuations and physiographic vicariance in shaping the distribution of southern African biodiversity.

Keywords: climate variability, *Gerbilliscus*, historical biogeography, megadroughts, mito-nuclear discordance, palaeodistributional modelling, southern Africa

Received 29 September 2014; revision received 27 August 2015; accepted 1 September 2015

Introduction

Studies utilizing multiple genetic markers abound in the phylogeographic literature, yet certain geographic regions remain little studied using such techniques,

Correspondence: Molly M. McDonough, Fax: (202) 633 0182; E-mail: McDonoughM@si.edu

thereby limiting our ability to infer pattern and processes for much of the earth's biodiversity (Beheregaray 2008). One such example, south-central Africa (following Cotterill 2003), contains a heterogeneous environment consisting of grassland, woodland and savanna–woodland habitats centred in the Zambezi River basin and its tributaries. In their review of phylogenetic patterns for nineteen savanna ungulates distributed across sub-Saharan Africa, Lorenzen *et al.* (2012) found surprisingly little phylogenetic structure for Pleistocene taxa distributed in southern Africa compared with populations in East Africa, and suggest that environmental stability and a long-standing savanna refuge in southern Africa conserved populations/species over time. However, given the relatively long generation times and greater dispersal capabilities of ungulates, they are not ideal candidates for examining fine-scale phylogeographic patterns. For smaller mammals in African savanna ecosystems, several studies in the Sudano-Sahelian savanna have provided insight into the evolutionary processes (e.g. vicariance and population expansion due to climatic cycles, allopatric divergence due to the presence of geographic barriers, and secondary contact between divergent lineages) that have shaped the geographic distributions of West African taxa (Brouat *et al.* 2009; Bryja *et al.* 2010; Granjon *et al.* 2012; Dobigny *et al.* 2013). For small mammals in southern Africa, the effects of historical processes for species occurring across large geographic areas have not yet been realized.

Fortunately, owing to anthropological and palaeontological research, there exists a rich fossil record for African small mammals (Denys 1999; Werdelin & Sanders 2010) and an advanced understanding of the palaeoclimatic conditions over the last 5 million years for eastern and southern Africa (deMenocal 2004; Feakins *et al.* 2005; Trauth *et al.* 2010). A recent synthesis of empirical palaeoclimate data hypothesized that the climatic variability that characterized East Africa during the Plio-Pleistocene played a key role in speciation and extinction in hominin lineages (Potts 2013). Although evolutionary patterns during these periods are seemingly less understood for other mammals, climate fluctuations during the Pleistocene have been implicated in faunal turnover, principally migration and extinction, for ungulates in the Cape Floral Region of South Africa (Faith & Behrensmeyer 2013 and references therein). While many African species likely endured through numerous climatic oscillations during the Pleistocene, there were time periods that were exceptional. Of particular interest were multiple intervals of extremely arid conditions, that is 'megadroughts', in Africa between 0.135 and 0.090 mya that are predicted to have been more influential on vegetation than the last glacial maximum (Cohen *et al.* 2007; Scholz *et al.* 2007).

Geographic distributions and genetic patterns for extant mammals suggest that climate change alone does not explain the vicariant distributions observed in southern Africa (Grubb *et al.* 1999; Cotterill 2003; Van Daele *et al.* 2007). For example, geomorphological changes, particularly in the course of the Zambezi River and its tributaries, have resulted in genetic divergence in populations of reduncine antelopes (Cotterill 2003) and subterranean rodents (Van Daele *et al.* 2004, 2007). Vicariant patterns such as these are expected given that southern African river systems during the Pliocene and Pleistocene were dynamic (Moore *et al.* 2007, 2012). In particular, the Zambezi – the largest river in southern Africa, experienced large fluctuations in its course within the last five million years, forming the modern course only within the last million years (1.1–0.65 Ma; Moore *et al.* 2012). Models of drainage evolution in this region indicate that the Palaeo-Chambeshi River, up until the Early Pleistocene, flowed from its headwaters in Tanzania to the southwest into the Kalahari basin and ultimately formed the current modern course of the Kafue River (Cotterill 2003; Moore *et al.* 2012). Catchments of ancient rivers in this region resulted in novel rivers and formation of wetlands, including a series of floodplains in the current Zambia–Zimbabwe–Botswana borders. It is predicted that the parapatric distributions found in many vertebrates in central Zambia resulted from bisection of the Palaeo-Chambeshi River (Ansell 1978; Cotterill 2003). While Cotterill (2003) argues that evidence for geomorphological effects is 'subtle and challenging to discover and decipher', a multilocus approach using a widespread taxon should illuminate signals of vicariance due to rivers or a number of topographical features resulting from tectonic processes (e.g. escarpments, mountains and plateaus).

Gerbilliscus (formerly *Tatera*) is a widespread rodent genus occurring throughout sub-Saharan Africa that includes three well-supported clades corresponding to eastern, western and southern geographic lineages, respectively, although recent evidence indicates paraphyly with respect to *Gerbillurus* (Colangelo *et al.* 2007; Granjon *et al.* 2012). Within the southern clade, the bushveld gerbil, *Gerbilliscus leucogaster,* is the most widespread species ranging from Tanzania to the northern half of South Africa (Dempster 2013) where it occurs in a variety of relatively dry habitats ranging from open grassland, wooded savanna to woodlands typically characterized by sandy soils [e.g. Miombo, Mopane, Zambezian; Dempster (2013)]. Musser & Carleton (2005) recognize 24 synonyms for *G. leucogaster*, indicating that geographical variation was historically recognized, a

hypothesis yet to be tested using genetic data. Given its widespread distribution, short generation time and high mitochondrial mutation rate (Granjon *et al.* 2012) compared with larger mammals, *G. leucogaster* provides an ideal candidate for examining potential geographic barriers to gene flow as well as effects of past climate change in southern Africa.

Herein, we use this broadly distributed rodent to test the hypothesis proposed by Lorenzen *et al.* (2012) that southern Africa acted as a long-standing refugium for savanna fauna. Given the potential for expansion and contraction of suitable habitat associated with Pleistocene climate oscillations, and in particular, the extremely arid periods known as megadroughts (e.g. those occurring between 0.135 and 0.090 mya), we hypothesize that habitat fragmentation resulted in refugial isolation of this widely distributed species during this time frame. We predict that signatures of regional isolation and expansion will be present in mitochondrial, autosomal and sex-linked data sets for *G. leucogaster*. In addition, we use ENM to examine changes in the distribution of abiotically suitable conditions for *G. leucogaster* in southern Africa during both interglacial and glacial maximum periods. Potential refugia identified with ENMs are compared to phylogeographic patterns derived from multiple genetic transmission elements to look for congruent patterns of isolation and expansion between these two data sets. We then use this combined approach to examine temporally congruent patterns between genetic signatures of demographic expansion and increases in suitable abiotic conditions during the Pleistocene. Finally, we use spatial diffusion analysis to determine whether geographic locations of the clade ancestors were present during times of refugial isolation estimated using ENM.

In addition to examining influences of Pleistocene climate change on the evolutionary history of *G. leucogaster*, we test whether genetic patterns for this species are associated with geomorphological changes, especially large rivers. For this purpose, we focused our sampling on individuals from opposite sides of major river systems in southern Africa. In particular, we focus our effort in central Zambia where the Kafue, Zambezi and Chambeshi rivers have undergone major changes in their course during the Pleistocene and are predicted to have influenced the distributions of many vertebrate taxa. Using a widespread savanna–woodland species, this study provides one of the first comparative assessments of the impacts of climate oscillations and geomorphological barriers to examine phylogeographic patterns for small mammals distributed across southern Africa.

Materials and methods

Taxon sampling and laboratory methods

Genetic data from recent collection trips or mitochondrial sequences deposited on GenBank were obtained from 91 individuals from 31 localities across the known distribution of *Gerbilliscus leucogaster* (Fig. 1A, Appendix S1, Supporting information). Genomic DNA was extracted using a DNeasy Blood and Tissue Kit (Qiagen Inc., Valencia, CA, USA). Mitochondrial cytochrome-*b* (Cytb), Y-linked UTY11 and X-linked TKTL1 were amplified and Sanger sequenced (Appendix S2, Supporting information). DNA sequences were aligned using SEQUENCHER version 4.9 (Gene Codes Corporation, Ann Arbor, MI, USA). For thirteen females included in the X chromosome data set (of 42 individuals), the heterozygous positions were scored as ambiguities. Number of mitochondrial haplotypes, haplotype diversity, nucleotide diversity and average number of nucleotide differences were calculated using DNASP version 5.0 (Librado & Rozas 2009). AFLPs were generated following the methods outlined in McDonough *et al.* (2008) (Appendix S2, Supporting information).

Phylogenetic estimates

Mitochondrial phylogenies from haplotypic data were estimated using Bayesian inference (BI) and maximum likelihood (ML). The best-fit model of evolution was determined by examining both AICc and BIC scores in JMODELTEST version 2.1 (Darriba *et al.* 2012). Bayesian inference was performed using MRBAYES version 3.2.1 (Ronquist *et al.* 2012). Four independent MCMC chains were run for 1×10^6 generations, which resulted in split frequencies <0.01. Trees were logged every 1000th iteration and 10% of trees were discarded as burn-in. Maximum-likelihood phylogeny was estimated using the online server for PHYML version 3.0 (Guindon *et al.* 2010) with a BIONJ starting tree (Gascuel 1997) and 1000 bootstrap replicates. Out-group taxa included other *Gerbilliscus* species from southern, eastern and western African clades as well as other gerbil genera (Appendix S1, Supporting information). Uncorrected p and Kimura 2-parameter genetic distances were calculated in MEGA version 5.0 (Tamura *et al.* 2011). To characterize interior-tip relationships among haplotypes for the mitochondrial, X and Y data sets, we used the median-joining network algorithm (Bandlet *et al.* 1999) which allowed for multistate data observed in the X chromosome marker. Haplotypes were estimated using the software POPART version 1 (http://popart.otago.ac.nz).

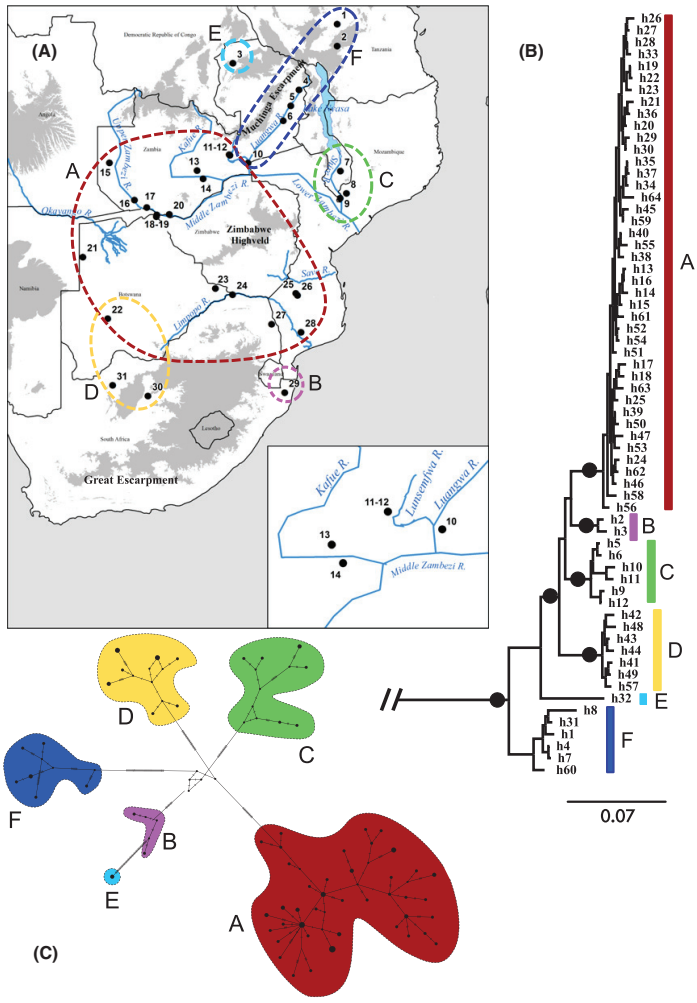


Fig. 1 Mitochondrial relationships of *Gerbilliscus leucogaster*. (A) Distribution of mitochondrial clades in southern Africa. Numbered dots indicate sampling localities. Inset emphasizes the Kafue Flats region and Kafue, Lunsemfwa and Luangwa rivers. Grey shading indicates elevation >1300 m. (B) Bayesian phylogeny with Bayesian posterior probabilities ≥0.95 and ML likelihood bootstrap ≥80% indicated by closed circles. (C) Median-joining network, branch lengths indicate evolutionary change.

Nuclear AFLP data set

A subset of individuals included in the mitochondrial data set was assessed for nuclear variation using AFLPs. To test our hypothesis that rivers acted as barriers to gene flow, genetic sampling effort was focused on individuals on alternate sides of rivers. GENALEX version 6 (Peakall & Smouse 2006) was used to generate the binary matrix of AFLP fragments, summary statistics, Nei's genetic distances (Nei 1972) and a principal coordinate analysis (PCoA). To estimate phylogenetic relationships, Bayesian inference was conducted in MRBAYES version 3.2.1 using the restriction site model allowing for coding bias and the option 'no-absence-sites' corresponding to the model for AFLP data sets proposed by Ronquist *et al.* (2005). The Dirichlet prior for the state frequencies was set to 0.4656 and 0.5344, corresponding to the ratio of ones to zeros in the original binary matrix. Four independent Markov chain Monte Carlo (MCMC) chains (three heated and one cold) were run for 100 million iterations. Trees were sampled every 1000th iteration and the first 10% was discarded as burn-in.

To assign individuals to genetic clusters (where K = the number of clusters), the MCMC clustering method in the program STRUCTURE version 2.3.2 was employed (Pritchard *et al.* 2000; Falush *et al.* 2003). To accommodate the dominant nature of AFLP data, the recessive allele model was used following Falush *et al.*

(2007). The recessive allele model treats the AFLP bands as partial information that is included in the Bayesian update used to estimate the probability of all possible genotypes (Falush *et al.* 2007). The admixture model with correlated allele frequencies was selected (Falush *et al.* 2003). Prior population assignment information was not included in the analysis. Values of *K* ranging from 1 to 10 were analysed with 10 independent simulations for each K and 600 000 total iterations with a burn-in of 100 000. To determine the optimal *K* value, STRUCTURE HARVESTER (Earl & vonHoldt 2012) was used to generate plots for Δ*K* (Evanno *et al.* 2005). The *Greedy* clustering algorithm in CLUMPP version 1.1.2 (Jakobsson & Rosenberg 2007) was used to align the 10 replicates, and the final visualization of the population clusters was conducted using DISTRUCT version 1.1 (Rosenberg 2004).

Timescale of diversification estimates

Divergence time estimates were calculated using the Cytb haplotype data set in BEAST version 1.7.4 (Drummond *et al.* 2012). The HKY+I+G nucleotide substitution model was included in the analysis, and codon positions were unlinked. Bayes factor tests performed following Suchard *et al.* (2001) for the various coalescent and speciation tree priors available in BEAST version 1.7.4 indicated the birth and death speciation prior following Gernhard (2008) was the best-fitting model for this data set. All likelihood ratio tests for a molecular clock (Felsenstein 1988) performed in MEGA version 5 (Tamura *et al.* 2011) were rejected; therefore, uncorrelated log-normal relaxed clock was employed.

Various out-group taxa were included (Appendix S1, Supporting information) to place fossil calibrations at crown nodes. Out-groups included East, West and southern African *Gerbilliscus*, *Gerbillurus*, *Tatera* and *Desmodillus*. Three internal fossil priors (using exponential distributions) were included that served as minimum dates for the common ancestor of that group/clade. For the crown height of all *Gerbilliscus,* an offset of 6 Ma (mean = 1.52) was used, corresponding to the oldest fossil for this genus from the Middle Awash, Ethiopia (Wesselman *et al.* 2009). The oldest fossils for the southern African *Gerbilliscus* clade from the Jägerquelle and Nosib fossil sites in Namibia (Senut *et al.* 1992) date to ~2.5 Ma. The first appearance date in the fossil record for *Gerbillurus* is also found at these Namibia sites (Denys 1999). For these clades (southern African clade and *Gerbillurus* clade), an offset = 2.5 and mean = 0.77 were used, which put the 97.5% quantile at 5.332, or the beginning of the Late Pliocene. Finally, tree height was set using a normal distribution with initial value set to 9.19 (mean = 9.19, SD = 1.62), which corresponds to the time to most recent common ancestor (TMRCA) of *Tatera* + *Desmodillus* + *Gerbilliscus* following the fossil calibrated molecular phylogeny of Chevret & Dobigny (2005).

For all analyses, chain length was set to 50 million iterations with data logged every 1000th iteration. TREE ANNOTATOR version 1.7.4 (Drummond *et al.* 2012) was used to summarize the posterior sample of trees, after removal of the first 10% of trees as burn-in, as a maximum clade credibility tree. TRACER version 1.5 (Rambaut & Drummond 2007) was used to evaluate the adequacy of the 10% burn-in and to summarize the posterior sampling of parameters. FIGTREE version 1.3.1 (Rambaut 2009) was used for visualization.

Demographic analyses

A representative of each of the three classes of demographic statistics outlined in Ramírez-Soriano *et al.* (2008) was used to estimate historical demographic parameters of ancestral clades and to identify patterns of population expansion. Mismatch distributions for each clade were estimated in ARLEQUIN version 3.11 (Excoffier *et al.* 2005) under the sudden expansion model using uncorrected pairwise differences, with significance determined using 1000 randomizations (Rogers & Harpending 1992). For distributions with nonsignificant raggedness indices, expansion time was estimated following Schenekar & Weiss (2011). Fu's *Fs* (Fu 1997) and Tajima's *D* (Tajima 1989) were also calculated, and negative and significant values of each of these statistics indicate recent population expansion (see Fahey *et al.* 2014). Significance was assessed by comparison of observed data to expectations under neutrality.

To complement mismatch distributions and estimate timing of demographic events, Extended Bayesian Skyline Plot (EBSP, Heled & Drummond 2008) was implemented in BEAST version 1.7.4 (Drummond *et al.* 2012) on clades that exhibited signatures of population expansion in the previous analysis. The EBSP included mitochondrial and sex chromosome data to estimate population change through time using multiple unlinked loci. Mutation rate calculated from the previous BEAST analysis was used to inform the prior distribution in the EBSP analysis. For the intraspecific mitochondrial data set, mutation rate and a strict clock model were applied including a normal prior with mean = 0.047 (95% HPDs = 0.0390–0.0551). Likewise, the Y chromosome marker was given a prior with mean = 0.004 (95% HPDs = 0.0024–0.0057) and X chromosome marker with mean = 0.003 (95% HPDs = 0.0017–0.0045), using prior mutation rates from sex chromosome data sets for a larger gerbil phylogeny (McDonough 2013). The analysis consisted of 50 million iterations sampled every 5000th iteration.

Ecological niche modelling

To provide an additional and independent test of the Pleistocene refugia hypothesis, ecological niche modelling (ENM) was performed under the Biotic, Abiotic and Movement conceptual framework (BAM; Soberón & Peterson 2005). We obtained 67 unique localities of *G. leucogaster* that included 31 freshly collected individuals used in molecular analyses and 36 historic museum specimens identified using a suite of diagnostic cranial measurements (Appendix S3, Supporting information). To reduce potential bias associated with clustered sampling localities, we spatially thinned occurrence data by selecting a random subset of <67 points with a minimum distance between points of 0.58 decimal degrees in ARCGIS 10.1 (ESRI, Redland, CA, USA). We performed this spatial thinning three times and used each subset to generate its own unique ENM. We used the 19 bioclimatic variables from WORLDCLIM (Hijmans *et al.* 2005; www.worldclim.org) at a spatial resolution of 2.5′ (~5 × 5 km) to predict the present (1950–2000) distribution of abiotically suitable conditions. This coarser resolution was chosen over the available 30 arc second data to better match the coordinate uncertainties associated with georeferenced, textual localities of museum specimens. Although colinearity between climatic variables has been highlighted as a serious concern in ecology (Dormann *et al.* 2013), we chose to include all 19 bioclimatic variables rather than an arbitrary subset of them for several reasons. First, although identifying correlations between variables is relatively straightforward, deciding which of the two correlated variables to include represents an arbitrary task, especially in the absence of detailed physiological limitations for a species. Second, although the application of ordination techniques such as principal component analysis can eliminate the need for arbitrarily dropping one of two correlated variables through the creation of latent variables, application of latent variables developed under one climate scenario cannot be used to project into alternative climate scenarios (Braunisch *et al.* 2013). Finally, machine learning algorithms such as those employed in MAXENT (Phillips *et al.* 2006) allow for correlated variables to be considered separately and interactively using nonlinear relationships, making them more suitable for handling correlated predictor variables than other modelling techniques (Braunisch *et al.* 2013). For the last glacial maximum (LGM, ~21 000 years BP) and last interglacial (LIG, ~120 000–140 000 years BP; Otto-Bliesner *et al.* 2006), we used the same 19 bioclimatic variables reprojected under the 'Community Climate System Model' (CCSM) circulation model (www.worldclim.org).

The regional extent in which models are calibrated (i.e. 'M') is known to impact ecological niche predictions in a pervasive manner (Anderson & Raza 2010; Barve *et al.* 2011), thus defining explicit and biogeographically relevant calibration areas is critical for appropriate model development and validation (Owens *et al.* 2013). We compared model results using four different 'M' designations: the current geographic distribution according to the International Union for Conservation of Nature (IUCN; www.iucn.org), the IUCN distribution buffered by 500 km, 1000 km and 2000 km (Fig. S1, Supporting information). ENMs under these four scenarios were generated using the maximum entropy algorithm (Phillips *et al.* 2006) in the program MAXENT 3.3.3k for each spatially thinned subset of localities (with 52, 50 and 51 localities for subsets 1, 2 and 3, respectively). We selected the autofeatures option in MAXENT, which allows for linear, quadratic, product, threshold and hinge features to describe relationships between specimen locations and environmental conditions (Merow *et al.* 2013). Given the potential risk of generating nonbiologically meaningful relationships, both clamping and extrapolating options were deselected for all models (Owens *et al.* 2013). We used raw outputs to avoid potential bias resulting from associations related to postprocessing techniques and the value of τ (Merow *et al.* 2013). A total of 10 bootstrap model replicates were generated for each subset, and the median value of these 10 models was used to calculate a grand median value among all three subsets. Model performance was evaluated using the area under the curve (AUC) with an 80 – 20% training–test data split. A jackknife test of variable importance was used to assess individual variable contributions to model predictions, although such contributions must be interpreted cautiously when using correlated variables such as all 19 bioclim products (Phillips *et al.* 2006). All raster processing and assessments were performed using a combination of ARCGIS 10.1 and the R statistical software environment (http://www.r-project.org/).

Spatial diffusion modelling using continuous phylogeography

To test the hypothesis of refugial isolation, Bayesian implementation of the spatial diffusion approach was used to infer geographic origins of mitochondrial clades at specific time slices during the Pleistocene. Often used for tracking viral epidemics, this method has also been applied to vertebrate systems to examine patterns of geographic expansion and to infer geographic locations of potential refugia (Chiari *et al.* 2012; Escobar García *et al.* 2012; Barlow *et al.* 2013; Camargo *et al.* 2013; Nascimento *et al.* 2013; Barnett *et al.* 2014). Unique mitochondrial haplotypes and corresponding geographic

coordinates were used to generate the.xml file in BEAUTI version 1.7.4. We performed BEAST continuous phylogeography, including applying random jitter (0.02 × 0.02 decimal degrees) to create unique coordinates for individuals from identical localities. The simple constant population size coalescent Gaussian Markov random field prior (Bayesian Skyride) was used as the demographic prior (Minnin *et al.* 2008). Rate of diffusion was allowed to vary across branches by applying a Cauchy Relaxed Random Walk (RRW) model (Lemey *et al.* 2010). MCMC chains were run for 100 million generations, sampling every 10 000 generations. TREE ANNOTATOR version 1.7.4 (Drummond *et al.* 2012) was used to summarize the posterior sample of trees including a 10% burn-in and to create a maximum clade credibility tree. TRACER version 1.5 (Rambaut & Drummond 2007) was used to evaluate the adequacy of the 10% burn-in and to summarize the posterior sampling of each parameter. SPREAD version 1.0.6 (Bielejec *et al.* 2011) was used to generate the keyhole mark-up language (kml) file for visualizing spatial diffusion using GOOGLE EARTH (2011). To examine spatial diffusion over time, the TimeSlicer option in SPREAD was used to slice through each of the 10 000 posterior trees at specific time intervals to create polygons with 80% highest posterior densities (HPD). To examine geographic changes across a wide time span, slices were taken at intervals 0.60 and 0.30 mya. Additionally, we compared time slices during the last interglacial (0.13 mya) and last glacial maximum (0.02 mya).

Results

Geographic distributions of mitochondrial haplogroups

Phylogenetic trees (BI and ML) and a haplotype network demonstrate support for six major haplogroups that correspond to distinct geographic regions (Fig. 1). The most widespread haplogroup, subsequently referred to as clade A, occurs throughout the central portion of southern Africa, extending from central Zambia southward to northern South Africa and from western border of Botswana eastward to Mozambique. Haplogroups B-F are isolated to relatively smaller geographic regions compared to haplogroup A. The number of haplotypes within clades ranged from one in clade E to 43 in the widespread clade A (Table 1). The haplotype diversity ranged from 0.89 in clade D to 0.98 in clade A (excluding clade B with only 2 sequences). Clade A exhibited the lowest nucleotide diversity (0.547% ± 0.000) and lowest number of average nucleotide differences (5.807). Nucleotide diversity and average number of nucleotide differences were highest in clades C and F (Table 1). Regions where mitochondrial clades occur in sympatry or parapatry exist in the Kalahari region of Botswana at site 22 where clades A and D occur in syntopy and in central Zambia where divergent clades A (sites 11 and 12) and F (site 10) are parapatric (Fig. 1A). The phylogenetic tree revealed a well-supported 'southern' group that includes clades A, B, C, D (Fig. 1B). Genetic divergence between each of the mitochondrial clades was relatively high, ranging from 3.5% to 6.8% (Table 2).

Nuclear data sets

The two-dimensional PCoA plot and a Bayesian unrooted phylogeny based on AFLP data revealed three major groups that corresponded to northern, central and southern geographic regions of southern Africa (Fig. 2). The northern cluster contained individuals from mitochondrial clades C, E, F and several individuals with mitochondrial A haplotypes that occur on the left bank of the Kafue and Lower Zambezi rivers, that is in the geographic region where mitochondrial clades A and F are parapatric in central Zambia (Fig. 2A–C; sites 11–13). Notably, individuals from site 14 (just opposite, right bank of the Kafue River) clustered with

Table 1 Demographic summary statistics for mitochondrial clades of *Gerbilliscus leucogaster*

Clade	*Ns*	*Np*	*Nh*	*Hd*	π (%)	*k*	Fu's *Fs*		Tajima's *D*	
A	61	78	43	0.984 ± 0.000	0.547 ± 0.000	5.807	−25.319	**	−2.241	**
B	2	7	2	1.000 ± 0.250	0.660 ± 0.000	7	—	—	—	—
C	7	26	6	0.952 ± 0.009	1.123 ± 0.000	11.714	0.59	ns	0.074	ns
D	12	20	7	0.894 ± 0.004	0.634 ± 0.000	6.727	0.069	ns	0.069	ns
E	2	—	1	—	—	—	—	—	—	—
F	7	39	6	0.952 ± 0.009	1.158 ± 0.000	12.286	−1.312	ns	−1.312	ns
All	91	199	64	0.990 ± 0.000	2.847 ± 0.001	29.698	—	—	—	—

Number of sequences (*Ns*), number of polymorphic sites (*Np*), number of haplotypes (*Nh*), haplotype diversity (*Hd*), nucleotide diversity (π) expressed as a percentage and average number of nucleotide differences (*k*), ** indicate values for Fu's *Fs* and Tajima's *D* that $P < 0.01$.

Table 2 Average uncorrected *p* (above) and Kimura 2-parameter (below) genetic distances between and within (grey-shaded) mitochondrial clades of *Gerbilliscus leucogaster*

Clade	A (*N* = 61)	B (*N* = 2)	C (*N* = 7)	D (*N* = 12)	E (*N* = 2)	F (*N* = 7)
A	0.008	0.035	0.041	0.043	0.063	0.061
B	0.036	0.006	0.037	0.044	0.049	0.051
C	0.043	0.038	0.012	0.046	0.061	0.055
D	0.045	0.046	0.048	0.007	0.062	0.062
E	0.068	0.052	0.065	0.066	0.000	0.057
F	0.065	0.054	0.058	0.066	0.060	0.008

the central group rather than the northern group. Remaining individuals in the widespread mitochondrial clade A represented the central group, although no individuals from clade B were included in the AFLP analysis. Genetic structure was present for individuals from sites north (14, 17 and 20) and south (15, 16, 18, 19 and 21) of the Zambezi River (Figs 2B and S2, Supporting information). Sites near the Limpopo and Save rivers (sites 23–27) cluster also in the central group (Figs 2 and S2, Supporting information). The southern group included individuals of the mitochondrial clade D from the xeric Kalahari region and individuals from mitochondrial clade A from site 22, where these two mtDNA lineages are sympatric in central Botswana (Fig. 2).

Plots of ΔK and ln $P(K)$ (Evanno *et al*. 2005) support the presence of two clusters in the STRUCTURE analysis, but we also include results for $K = 3$ as suggested by the PCoA and Bayesian phylogeny. $K = 2$ demonstrates distinct northern and southern groups with the northern cluster containing individuals in clades C, E and F as well as individuals in clade A that occur in Zambia on the left bank of the Kafue and Lower Zambezi (Fig. 3; sites 11–13). The southern cluster contained most individuals with A and all with D haplotypes. Clusters of $K = 3$ reveal the northern (grey), central (white) and southern (dark grey) groups exhibited also in the PCoA. Evidence of admixture between central and southern groups is also present in individuals from sites 23–27 (Fig. 3), indicating the Limpopo River is not a barrier to gene flow.

Phylogeographic structure was not recovered in the X chromosome data set (Fig. S3A, Supporting information). Mitochondrial clade A individuals exhibited the most X chromosome allelic diversity, often with greater mutational distance among clade A individuals than between other clades. The Y chromosome haplotype network exhibited stronger haplogroup structure than the X chromosome data set despite lower overall polymorphism (Fig. S3B, Supporting information). Individuals from northern Zambia (light blue, mitochondrial haplogroup E) exhibited unique Y and X chromosome haplotypes. Additionally, the southern group (yellow, mitochondrial clade D) shared a unique Y haplotype. Two male individuals (TK172838 and TK170589) with mitochondrial haplotype A displayed a Y chromosome haplotype similar to mitochondrial haplotype D individuals from the same geographic locality (site 22); this result is consistent with the nuclear AFLP data set (Fig. S3, Supporting information).

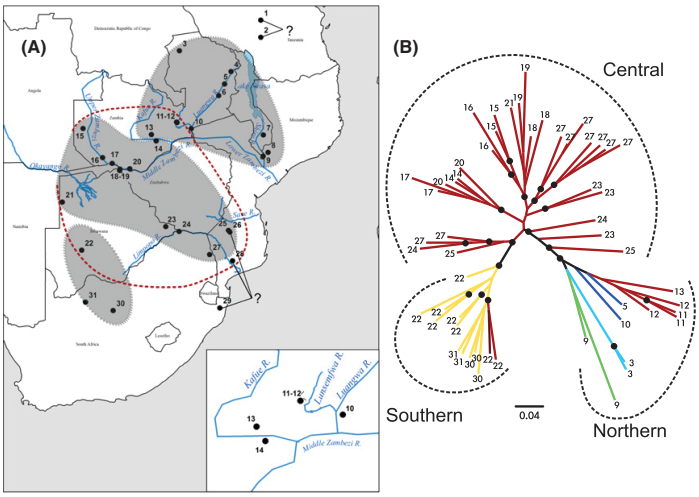
Divergence times, demographics and population expansion estimates

The TMRCA for all *Gerbilliscus leucogaster* clades was estimated to the mid-Pleistocene (1.046 mya; 95% HPDs = 1.415–0.741 mya) (Table 3). Individual mitochondrial clades diverged between 0.268 and 0.100 mya (HPDs = 0.408–0.35 mya).

Negative and significant values of Fu's *Fs* and Tajima's *D* were observed for clade A (Table 1). Based on raggedness results of the mismatch distribution, time since population expansion calculated for clade A was estimated to be approximately 0.060 mya (Fig. 4, Table S1, Supporting information). EBSP analysis (Fig. 4) indicated population expansion for clade A around 0.050 mya (95% HPDs 0.065–0.035 mya).

Ecological niche modelling

Models based on the four different areas of extent (i.e. M scenarios) performed qualitatively similar with comparative predictions (Figs S1 and S4, Supporting information). Given these results, and the fact that only the 2000-km buffer included most of East Africa, the known geographic origin of *Gerbilliscus*, all final models depicted were generated using the 2000-km buffer training area. The training data set from the 80% to 20% split present conditions model had a median AUC of 0.95 based on 10 replicates, indicating that model prediction accuracy was greater than random. In addition, the testing data set's median AUC was 0.93, supporting the model's predictive ability as well as its lack of overfitting.



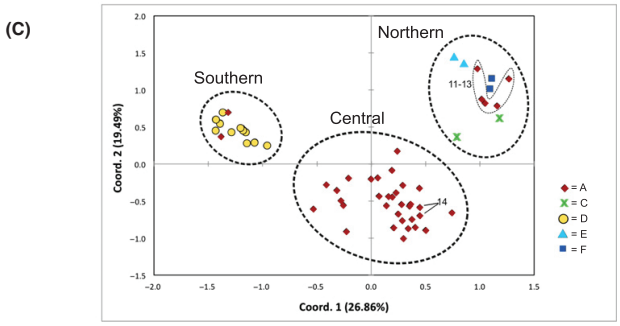


Fig. 2 Nuclear amplified fragment length polymorphism (AFLP) relationships of *Gerbilliscus leucogaster*. (A) Distribution of AFLP clusters in grey indicating northern, central and southern groups. Dashed line indicates boundaries of mitochondrial clade A. Inset emphasizes the Kafue Flats region and Kafue, Lunsemfwa and Luangwa rivers. '?'indicates localities not included in AFLP analysis. (B) Bayesian unrooted tree with localities at terminal branches, black closed circles indicate nodes with posterior probabilities ≥0.95. (C) Principal coordinate analysis (PCoA) ordination of northern, central and southern groups. Symbols and colours correspond to mitochondrial clades. Sites 11–14 are highlighted to reflect nearby geographic regions with similar mitochondrial haplotypes and divergent nuclear clusters.

Surprisingly, the LIG distribution of suitable conditions poorly reflected the distribution of currently suitable conditions. In fact, suitable conditions were mostly confined to west-central Angola and the southeastern coastal region of southern Africa (Fig. 5A). In addition, a large strip of highly suitable conditions starting in Ethiopia and stretching west along the southern boundary of the Sahel was recovered (Fig. S5, Supporting information).

The LGM distribution of suitable conditions appeared to reflect several of the phylogeographic breaks documented within *G. leucogaster* (Fig. 5B). Most notable was the large area of unsuitability corresponding to the Zimbabwe Highveld, which appears to support separation of the range into a northern and southern component, with the presence of small band of intermediate suitability linking the two along the Zimbabwe–Zambia border (Fig. 5B). Compared to the LIG, the extent of suitable conditions appears to have expanded during the LGM; however, areas of suitable conditions during the LGM were less extensive than those found in the present.

The present-day distribution of suitable abiotic conditions closely matched the current IUCN range map (Fig. 5C). The models identified centres of suitability that correspond with genetic lineages observed within *G. leucogaster*, concentrated in Botswana, northern South Africa, southern Mozambique and Zambia (Fig. 5C).

Spatial diffusion analysis

We summarized the dispersal process for *G. leucogaster* at four time slices (Fig. 6). By 0.6 mya, the 80% HPDs of the ancestral locations contained two major regions: a region in northern Zambia, northern Mozambique and southern Malawi, and a region in southern Zambia, the Kalahari arid region of Botswana, and South Africa (Fig. 6A). By 0.3 mya, most of the spread within *G. leucogaster* had occurred (Fig. 6B). The time slice during the LIG, which was predicted as highly unsuitable in the ENM data set, revealed that clades were present in geographic regions that could be indicative of potential refugia (Fig. 6C). By the LGM, additional maximum

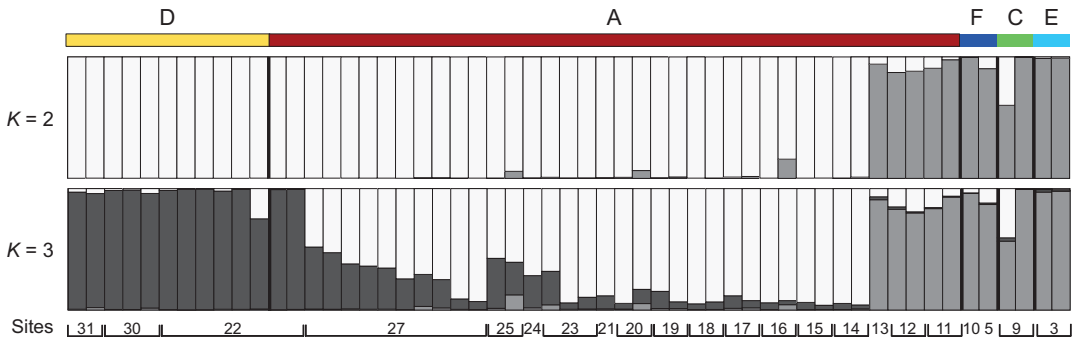


Fig. 3 Population clustering of *Gerbilliscus leucogaster* estimated from nuclear amplified fragment length polymorphism (AFLP) data based on $K = 2$ (top) or $K = 3$ (bottom). Individuals are represented as vertical bars shaded white, light grey or dark grey according to Bayesian population clustering. Mitochondrial clades depicted as coloured bars above individuals and geographic sampling numbered below.

Table 3 Estimates for time to most recent common ancestor (TMRCA) using unique haplotypes (clade E has only a single haplotype) for each of the mitochondrial clades and all *Gerbilliscus leucogaster* clades

Clade	TMRCA (Ma)	95% HPD
A	0.163	0.107–0.240
B	0.098	0.035–0.188
C	0.210	0.125–0.319
D	0.129	0.072–0.201
F	0.268	0.158–0.403
All	1.046	0.741–1.415

clade credibility (MCC) branches extended into the Kalahari arid regions of northern Botswana and South Africa (Fig. 6D).

Discussion

Demographic effects of Pleistocene climate variability in southern Africa

Our study highlights the phylogeographic history for a widespread rodent in southern Africa using a combination of multilocus molecular markers and palaeoecological niche modelling. Our results support the hypothesis that climate variability during the Late Pleistocene played a role in genetic divergence within *Gerbilliscus leucogaster*. These results do not correspond with previous hypotheses that southern Africa was a single stable, long-standing refugium of savanna-like habitats during the Pleistocene (Lorenzen *et al.* 2012; Maslin *et al.* 2012). Rather, we find that during the LIG (~0.140–0.120 mya), there was a strong reduction in suitable abiotic conditions for *G. leucogaster* (Fig. 5). Given that all of the clades (except F which is the oldest lineage) have 95% HPDs that coalesced during the LIG (Table 3), we find support that the reduction in suitable conditions during this time frame resulted in vicariance associated with refugial isolation.

Plio-Pleistocene climate change has long been implicated as a major factor contributing to East African hominin evolution, and more recently, consideration of the overall effects of climate instability during this time frame provides clues to why some lineages have persisted while others went extinct (Potts 2013). For example, the diversification of extant lineages of *G. leucogaster* occurred during the last prolonged period of substantial climate variability that occurred ~0.360–0.050 mya – a time frame encompassing a known decline of diversity of East African grazing mammals in a two-step process – first in nonbovid grazers followed by bovids (Potts 2013). Potts (2013) hypothesized that the taxa that were able to persist and disperse were those that were able to survive periods of climate variability. We propose that *G. leucogaster* provides a convincing example of a species that was able to tolerate the extreme and prolonged periods of instability characterizing the Late Pleistocene of southern Africa.

The multiple intervals of extremely arid conditions between 0.135 and 0.090 mya (Cohen *et al.* 2007; Scholz *et al.* 2007) likely substantiate the reduction in suitable abiotic conditions observed during the LIG. During these arid conditions, there is evidence of high pollen accumulation for Poaceae, suggesting opportunities for dry savanna expansion (Beuning *et al.* 2011). However, between 0.105 and 0.093 mya, there was a megadrought so severe that even xeric grassland expansion would have been limited. This event might have contributed to range retractions into refugia and therefore

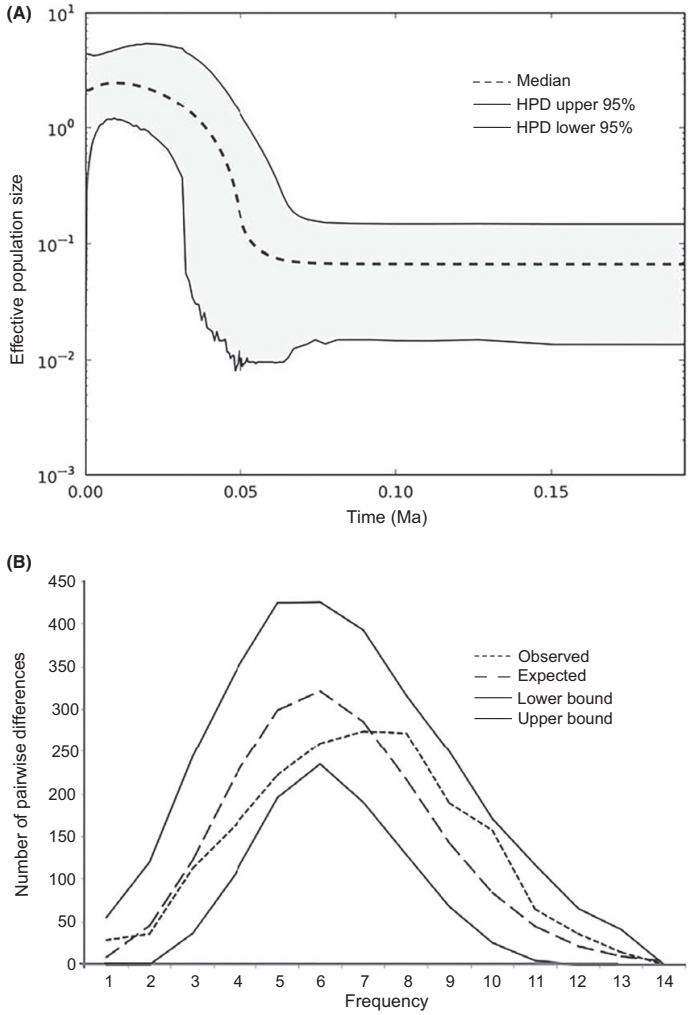


Fig. 4 Demographic signature of population expansion for *Gerbilliscus leucogaster* clade A. (A) Extended Bayesian Skyline Plot and (B) mismatch distribution estimated under the sudden expansion model.

opportunities for vicariance, as has been suggested for Zambezi Valley and Luangwa valley populations of bushbuck (Moodley & Bruford 2007). As *G. leucogaster* was widespread throughout the Zambezi region by this time (Fig. 6), this megadrought likely contributed to demographic changes within populations involving isolation of previously connected groups and subsequent differentiation through genetic drift. While southern African gerbils appear to tolerate a variety of arid and semi-arid habitats (Campbell *et al.* 2011), they are opportunistic breeders with population size dependent on food availability (i.e. seeds and arthropods; White *et al.* 1997; Blaum *et al.* 2006) and therefore are susceptible to environmental perturbations, which could potentially influence dispersal and colonization of new areas.

The LIG (0.140–0.120 mya) models for *G. leucogaster* recovered reduced areas of suitable conditions across southern Africa, with areas of suitability restricted to coastal and mid-interior regions of Mozambique and South Africa, although a single contiguous area of low suitability did appear in eastern Angola (Fig. 5). Barlow *et al.* (2013) predicted the presence of a similar coastal refugium for the African puff adder (*Bitis arietans*), however, their prediction was based on LGM conditions only. Although temporally discordant, the recovery of similar refugial conditions in this region likely explains the divergent coastal mitochondrial clade B in our data

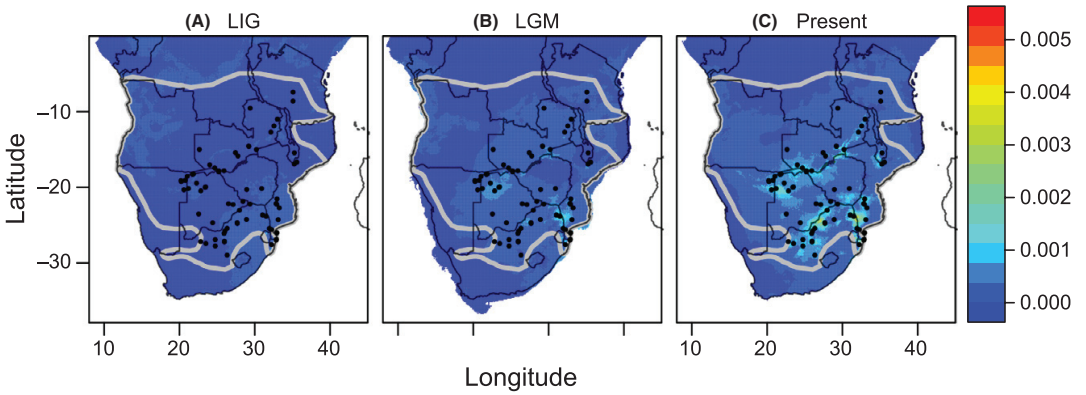


Fig. 5 Ecological niche models for (A) last interglacial, (B) last glacial maximum and (C) present-day conditions. Colours (grey to green) represent suitability of abiotic conditions and black dots depict locations of specimens used to generate models. The broad grey outline represents the current known range limit of *Gerbilliscus leucogaster* according to the IUCN.

set. In contrast, only small, isolated patches of low suitability were recovered during the LIG across Botswana, Zimbabwe and Zambia. In addition, the appearance of highly suitable conditions for *G. leucogaster* in the Horn of Africa, including Ethiopia, and along the southern edge of the Sahel appears inconsistent with temporal and spatial estimates of divergence patterns within *G. leucogaster* (Fig. S5, Supporting information). Although the Sahel region has no known fossil or modern records of *G. leucogaster*, other extant species of *Gerbilliscus* exist within this area. Given that our ENMs represent a Grinnellian niche, that is a combination of scenopoetic (noninteractive environmental conditions) variables which result in a positive intrinsic growth rate (Soberón 2007), ultimately ignoring Eltonian concepts of biotic interactions, it is not surprising that our model predicts areas of suitable conditions not currently occupied by *G. leucogaster*. In addition, a lack of comprehensive understanding of the accessibility of such areas to this species over an evolutionary timescale as well as the impacts of putative biotic interactions with other species of gerbils found in these regions ultimately prevents us from developing an accurate model of the true distribution ('P' of the BAM diagram; Soberón & Peterson 2005) for *G. leucogaster* using these methods. Despite these limitations, these continental-wide predictions (Fig. S5, Supporting information) provide useful data for generating future hypotheses regarding *Gerbilliscus* distribution patterns across sub-Saharan Africa.

Unfortunately, few studies incorporating ecological niche models for LIG conditions exist for southern Africa, limiting our abilities to contrast our patterns to those resulting from studies of other taxa. One of the few studies applying ENM in southern Africa used present-day climatic conditions as a proxy for interglacial climates and therefore did not provide LIG projections of present-day models (Barlow *et al.* 2013). Although present-day conditions fall within an interglacial period, a recent study of the Sahara found that climatic and hydrological cycles dramatically differed in LIG from those experienced today (Coulthard *et al.* 2013), highlighting potential discrepancies between present-day and LIG conditions for various regions of Africa. Similarly, our results do not support similar suitable conditions between present and LIG time periods, a pattern concordant with recent studies on the Pleistocene distribution of suitable conditions for the spotted hyena (*Crocuta crocuta*; Varela *et al.* 2009) and hominids (*Homo sapiens*; Drake *et al.* 2011). Taken in concert with our findings, it appears that assuming that the distribution of suitable environmental conditions during the LIG closely mimics those of modern time periods could be incorrect, especially for taxa from southern Africa. Such mismatches between current and projected models could result from several factors, including but not limited to (i) lack of appropriate representation of the entire climatic spectrum of suitable environmental conditions for the species (i.e. our species occurrence records only represent a subset of all inhabited localities; Varela *et al.* 2009), (ii) varying climatic relationships for correlated variables across time periods (Braunisch *et al.* 2013), and (iii) altered distributions of suitable conditions during the last interglacial [e.g. increased grassland coverage over much of North Africa during the LIG, (van Andel & Tzedakis 1996; Montoya 2007)]. Our results also highlight the risk of relying on a single time slice of climatic conditions to identify Pleistocene refugia for a particular species,

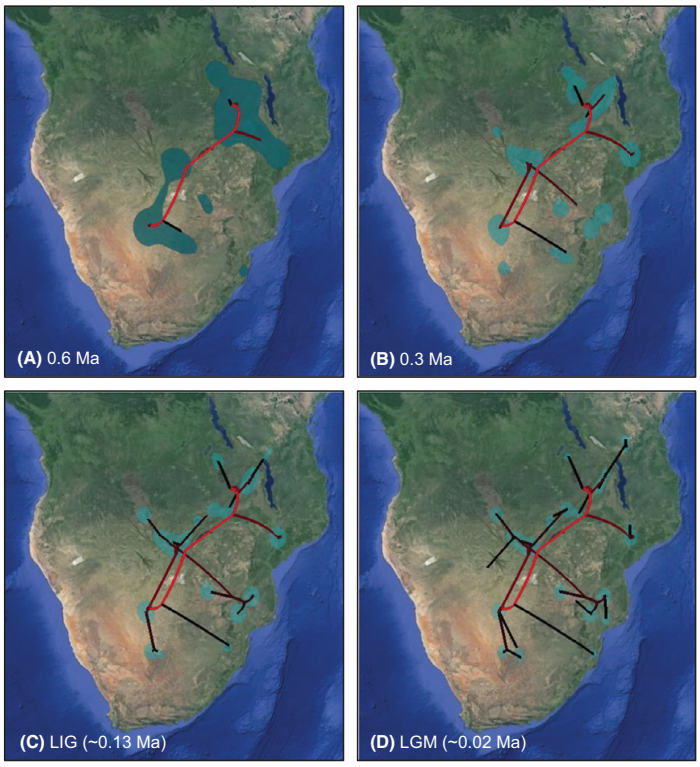


Fig. 6 Estimates of spatial diffusion at four time points during the Pleistocene (A) 0.6 mya, (B) 0.3 mya, (C) 0.13 mya (D) and 0.02 mya. Green coloured polygons indicate uncertainty (80% HPDs) surrounding geographic locations of internal nodes of the tree. Red-black colour gradient of the maximum clade credibility (MCC) tree branches informs the relative age of the transition (older–recent).

especially when that species exists in a region characterized by highly unstable and fluctuating climatic conditions, such as those found in southern Africa during the Pleistocene (Potts 2013).

Our data also provide evidence of population expansion for widespread mitochondrial clade A ~0.065–0.035 mya (Fig. 4), which corresponds with predicted expansion of suitable abiotic conditions after periods of prolonged drought. Periods of vegetation recovery followed the droughts, and by 0.085–0.080 mya, there was a resurgence of savanna–woodland habitat documented in the pollen core at Lake Malawi (Beuning *et al.* 2011). This was followed by another cooling event with temperatures decreasing from 26 to 20 °C between 0.068 and 0.060 mya (Woltering *et al.* 2011) which likely coincides with the signal of population expansion recovered for clade A. The last glacial maximum (0.025–0.015 mya) in this region is characterized by expansion of dry woodland and grassland habitat (Beuning *et al.* 2011). These results are consistent with population expansion of a similarly distributed clade of pouched mouse that underwent demographic expansion in the Zambezi region ~0.100–0.050 mya (Mikula *et al.* in review).

Influence of rivers on genetic structure

Our results also support the hypothesis that large rivers acted as barriers to gene flow for *G. leucogaster* populations. The most pronounced pattern observed in *G. leucogaster* was the genetic divergence among lineages in central Zambia. We found differing results between the mitochondrial versus nuclear data sets. For example, the mitochondrial data set revealed a statistically supported phylogenetic break between mitochondrial haplogroups A and F near the Lunsemfwa and Luangwa rivers, indicating a historical barrier in this region (Fig. 1). The phylogenetic break in the nuclear AFLP data occurs further south, indicating the Kafue River is currently a strong barrier to gene flow between populations. The Kafue River region is a flat landscape (i.e. Kafue Flats) that during the Pleistocene acted as a floodplain for this region and in fact supported Paleolake Patrick (now Lower Kafue River; Moore *et al.* 2012), which likely explains the strong phylogenetic signal.

The most convincing studies supporting the influence of the Zambezi River and its tributaries on speciation in mammals has been illustrated in genetic studies of

subterranean (Van Daele *et al.* 2004, 2007) and epigeic rodents (Mikula *et al.* in review), reduncine antelopes (Cotterill 2003) and the general distributional patterns across a broad array of mammals, including isolated endemic taxa (e.g. *Giraffa angolensis*, *Damaliscus lunatus*, among others) in this region (Ansell 1978; Grubb *et al.* 1999). For example, there is particularly high diversity of small mole-rats in the Zambezian region, including three chromosomal races/species of *Fukomys* that occur along a 250-km stretch of the Kafue River (Van Daele *et al.* 2004, 2007). This same geographic region (i.e. Kafue Flats) is where we found a strong phylogenetic break between the northern and central nuclear AFLP genetic groups in *G. leucogaster*. Further examples of isolation are found in divergent mole-rat lineages further to the north where restructuring of the Palaeo-Chambesi River occurred during the Pleistocene – and may explain the deep divergence exhibited in the isolated mitochondrial clade E. However, given our limited sampling regime, we cannot rule out that this clade could be more widely distributed, especially into southeastern Democratic Republic of Congo (DRC), which was predicted to be abiotically suitable in the ENM data set (Fig. 5A). The Katanga region in southeastern DRC is home to highly divergent lineages of mammals (e.g. *Heliophobius argentocinereus*; Faulkes *et al.* 2011), herpetofauna (Broadley & Cotterill 2004) and birds (Cotterill 2006). Unfortunately, genetic sequence data for *G. leucogaster* from the DRC are currently lacking.

Other geomorphological features, such as the Muchinga escarpment (adjacent to the Luangwa River, Fig. 1), have acted as long-term barriers to gene flow for some terrestrial mammals. We observed a distinct lineage (clade F) that occurs along a narrow strip of Zambezian and Mopane woodlands in western Tanzanian and southward throughout the Luangwa Valley Rift, a southward extension of the Great Rift Valley. This region contains endemics such as Thornicroft's giraffe (Fennessy *et al.* 2013), Cookson's Wildebeest, squirrels (Cotterill 2003), bushbuck (Moodley & Bruford 2007) and others (see Ansell 1978 and Cotterill 2003). Additionally, the Drakensberg Mountains on the eastern portion of the Great Escarpment likely acted as a barrier to dispersal for the geographically isolated mitochondrial clade C individuals in Mozambique, a pattern that is also observed in another savanna rodent, *Saccostomus* (Mikula *et al.* in review).

Patterns in mitochondrial, autosomal and sex chromosome-linked markers

By utilizing multiple genetic markers, we were able to characterize signatures of hybridization among mitochondrial clades and areas of admixture. Overall, we found less genetic structure in the nuclear AFLP data set compared to the mitochondrial data set. This mito-nuclear discordance was observed in geographic regions of contact between divergent mitochondrial clades A and D in central Botswana and between clades A and F in central Zambia. These regions of secondary contact in which introgression was inferred support the hypothesis that climate change promoted geographic isolation followed by subsequent population expansion. Patterns of variation at the Y chromosome marker describe shared haplotypes at geographic localities of admixture. Combined patterns are consistent with expectations under male-biased dispersal, but it should be noted that the results of demographic analyses also support an overall geographic expansion of clade A. Thus, observed patterns of admixture are likely a result of range expansion coupled with the general mammalian characteristic of males dispersing greater distances than females (see examples in Toews & Brelsford 2012). The X chromosome marker, which displayed greater haplotypic diversity than the Y chromosome marker, indicated complex patterns. This increased diversity, especially for clade A, is consistent with both influences of a larger effective population size relative to the Y chromosome, and an increased rate of haplotype origination and retention in the demographically expanding clade A. Influences of incomplete lineage sorting are indicated by X haplotypes distributed across all major geographic regions, a pattern not observed in the primarily autosomal AFLP data set, nor in the mitochondrial data set. Also, of note is the observation that mitochondrial clade E displayed distinct Y and X chromosomes and admixed AFLP loci, which may suggest that patterns of admixture with this region are reduced with respect to patterns at other sites where admixture was inferred.

Patterns for other mammalian systems in southern Africa

Phylogeographic interpretations for southern Africa mammals have been limited by studies based on restricted geographic sampling and emphasis on medium to large sized taxa with high dispersal capabilities. Furthermore, previous studies were often based on single-locus data sets (usually mitochondrial sequences) from which introgression and other evolutionary processes could not be inferred. For ungulates living in savannas, Lorenzen *et al.* (2012) found a general lack of phylogenetic structure and low genetic variation observed in southern populations compared to East African lineages. However, specific groups including the giraffe, *Giraffa camelopardalis*, do exhibit population genetic structure with three different subspecific forms

in East Africa and at least three lineages in southern Africa (Brown *et al.* 2007; Fennessy *et al.* 2013). Likewise, various *Papio* lineages radiated into southern Africa ~1.8 Ma and share a similarly complex phylogenetic history as *G. leucogaster* (Zinner *et al.* 2009). For example, three mitochondrial clades of *Papio* come together in a region of secondary contact along the Zambezi River basin near the Kafue and Luangwa rivers, suggesting patterns of divergence within this genus were also influenced by this river system.

Conflicting patterns are also observed for faster evolving species with lower dispersal rates. Phylogeographic estimation for the most widespread African rodent, the multimammate mouse, *Mastomys natalensis*, found little geographic structure within the southern African mitochondrial haplogroup (Colangelo *et al.* 2013; haplogroup IV); however, microsatellite analysis for the same populations indicates the Zambezi and Kafue rivers function as a strong barrier to gene flow (J. Zima, V. Mazoch, R. Šumbera & J. Bryja, in prep). For southern African savanna-adapted rodents, the genus *Rhabdomys* exhibits regional divergence including a distinct phylogenetic break between South Africa and Zimbabwe, likely the result of the Limpopo River (Castiglia *et al.* 2011). Additional mammalian phylogeographic studies restricted to smaller geographic regions, mainly focused in South Africa, reveal phylogenetic signals that correspond to lineage specific affinities to regional biomes (Russo *et al.* 2010; Edwards *et al.* 2011). Of particular relevance to our results, Russo *et al.* (2010) found that each of the eight lineages of the Namaqua rock mouse (*M. namaquensis*) had a strong association with specific biomes in southern Africa, including distributions that closely match those of *G. leucogaster*. For example, *M. namaquensis* 'lineage G' (Russo *et al.* 2010) is restricted to the Eastern Kalahari Bushveld (savanna) biome, which is consistent with the distribution found for *G. leucogaster* clade D. These results lend support to the presence of various ecological affinities in each of the *G. leucogaster* haplogroups – with haplogroup A and more northern haplogroups inhabiting various savanna–woodland habitats and the southern AFLP genetic group occurring in more xeric grassland habitats (see ecoregions in Appendix S1, Supporting information). Similar findings were reported in Africa's most widespread ungulate, the bushbuck, which has >20 lineages that diverged within the last million years that contain examples of both habitat specialists and generalists (Moodley & Bruford 2007).

Conclusions

Here, we show support for six divergent mitochondrial lineages across the geographic range of *Gerbilliscus leucogaster* that shared a common ancestor within the last million years. During this time frame, southern Africa was subject to climate oscillations shifting between warm-wet and cool-dry periods punctuated with intervals of extreme drought (i.e. megadroughts). We found evidence that population size and abiotic conditions for *G. leucogaster* were subject to expansion and contraction that likely coincided with fluctuations in savanna–woodland habitat during the Late Pleistocene. Additionally, we demonstrate that changes in the course of the Zambezi River and its tributaries influenced historic and current distributions for *G. leucogaster*. These results and other patterns elucidated herein provide strong evidence for evolutionary responses operating at different biological and phylogenetic scales, highlighting a complex history for *G. leucogaster* shaped by climate change, physiogeographic vicariance and subsequent demographic responses to such events. While our results indicate greater phylogeographic structure for southern Africa than previous hypotheses for savanna ungulates, this may be due to specific habitat/soil requirements for these rodents compared to larger mammals. Comparative phylogeography using data sets from a variety of taxa with differing life history strategies and habitat requirements will likely reveal deeper insight into the historical biogeography of the Zambezi region.

Acknowledgements

For assistance with Botswana permits, we thank T. Nkwane with the Botswana Ministry of Environment, Wildlife and Tourism [permit number: EWT 8/36/4 XVI (46)]. For supporting the fieldwork in Botswana, we thank the Botswana National Museum, including M. Tswiio and K. Lemson and especially N. Mosesane, J. Marais, D. Thomas of Berry Bush Farms, B. and M. Sieberhagen of Kalahari Rest and the village development committee chairperson at Lepokole Village. For help with collection of samples in Malawi, Zambia and Tanzania, we acknowledge J. Šklíba, M. Lövy, E. Hrouzková, C. Sabuni, G. Mhamphi, H. Konvičková and all local collaborators. The assistance of R. Makundi and W.N. Chitaukali with project logistics is highly appreciated. For permission to carry out the research and to collect specimens in Malawi, we are obliged to the National Research Council, in Tanzania to Sokoine University of Agriculture and in Zambia to Zambian Wildlife Authority. Phylogeographical research in East African savannas was supported by the Czech Science Foundation, projects no. P506-10-0983, to JB and RS and no. 15-20229S, to JB. Additional funding for fieldwork and laboratory work was supported by an Explorers Club Exploration Fund Research Grant, Texas Academy of Science Research Grant and American Society of Mammalogists grants-in-aid to MMM. For assistance with access to museum specimens housed in the National Museum of Natural History, we thank K. Helgen and D. Lunde.

References

van Andel TH, Tzedakis PC (1996) Palaeolithic landscapes of Europe and environs, 150,000–25,000 years ago: an overview. *Quaternary Science Reviews*, **15**, 481–500.

Anderson RP, Raza A (2010) The effect of the extent of the study region on GIS models of species geographic distributions and estimates of niche evolution: preliminary tests with montane rodents (genus *Nephelomys*) in Venezuela. *Journal of Biogeography*, **37**, 1378–1393.

Ansell WFH (1978) *The Mammals of Zambia*. National Parks and Wildlife Services, Chilanga, Zambia.

Bandlet HJ, Forster P, Röhl A (1999) Median-joining networks for inferring intraspecific phylogenies. *Molecular Biology and Evolution*, **16**, 37–48.

Barlow A, Baker K, Hendry CR *et al.* (2013) Phylogeography of the widespread African puff adder (*Bitis arietans*) reveals multiple Pleistocene refugia in southern Africa. *Molecular Ecology*, **22**, 1134–1157.

Barnett R, Yamaguchi N, Shapiro B *et al.* (2014) Revealing the maternal demographic history of *Panthera leo* using ancient DNA and a spatially explicit genealogical analysis. *BMC Evolutionary Biology*, **14**, 70.

Barve N, Barve V, Jiménez-Valverde A *et al.* (2011) The crucial role of the accessible area in ecological niche modeling and species distribution modeling. *Ecological Modelling*, **222**, 1810–1819.

Beheregaray LB (2008) Twenty years of phylogeography: the state of the field and the challenges for the Southern Hemisphere. *Molecular Ecology*, **17**, 3754–3774.

Beuning KRM, Zimmerman KA, Ivory SJ, Cohen AS (2011) Vegetation response to glacial–interglacial climate variability near Lake Malawi in the southern African tropics. *Palaeogeography, Palaeoclimatology, Palaeoecology*, **303**, 81–92.

Bielejec F, Rambaut A, Suchard MA, Lemey P (2011) SPREAD: spatial phylogenetic reconstruction of evolutionary dynamics. *Bioinformatics*, **27**, 2910–2912.

Blaum N, Rossmanith E, Jeltsch F (2006) Land use affects rodent communities in Kalahari savannah rangelands. *African Journal of Ecology*, **45**, 189–195.

Braunisch V, Coppes J, Arlettaz R, Suchant R, Schmid H, Bollmann K (2013) Selecting from correlated climate variables: a major source of uncertainty for predicting species distributions under climate change. *Ecography*, **36**, 971–983.

Broadley DG, Cotterill FPD (2004) The reptiles of southeast Katanga, an overlooked 'hot spot'. *African Journal of Herpetology*, **53**, 35–61.

Brouat C, Tatard C, Bâ K *et al.* (2009) Phylogeography of the Guinea multimammate mouse (*Mastomys erythroleucus*): a case study for Sahelian species in West Africa. *Journal of Biogeography*, **36**, 2237–2250.

Brown DM, Brenneman RA, Koepfli K-P *et al.* (2007) Extensive population genetic structure in the giraffe. *BMC Biology*, **5**, 57. doi:10.1186/1741-7007-5-57.

Bryja J, Granjon L, Dobigny G *et al.* (2010) Plio-Pleistocene history of West African Sudanian savanna and the phylogeography of the *Praomys daltoni* complex (Rodentia): the environment/geography/genetic interplay. *Molecular Ecology*, **19**, 4783–4799.

Camargo A, Werneck FP, Morando M, Sites JW Jr, Avila LJ (2013) Quaternary range and demographic expansion of *Liolaemus darwinii* (Squamata: Liolaemidae) in the Monte Desert of Central Argentina using Bayesian phylogeography and ecological niche modelling. *Molecular Ecology*, **22**, 4038–4054.

Campbell TL, Lewis PJ, Williams JK (2011) Analysis of the modern distribution of South African Gerbilliscus (Rodentia: Gerbillinae) with implications for Plio-Pleistocene palaeoenvironmental reconstruction. *South African Journal of Science*, **107**, 1–7.

Castiglia R, Solano E, Makundi RH, Hulselmans J, Verheyen E, Colangelo P (2011) Rapid chromosomal evolution in the mesic four-striped grass rat *Rhabdomys dilectus* (Rodentia, Muridae) revealed by mtDNA phylogeographic analysis. *Journal of Zoological Systematics and Evolutionary Research*, **50**, 165–172.

Chevret P, Dobigny G (2005) Systematics and evolution of the subfamily Gerbillinae (Mammalia, Rodentia, Muridae). *Molecular Phylogenetics and Evolution*, **35**, 674–688.

Chiari Y, van der Meijden A, Mucedda M, Lourenço JM, Hochkirch A, Veith M (2012) Phylogeography of Sardinian cave salamanders (genus *Hydromantes*) is mainly determined by geomorphology. *PLoS ONE*, **7**, e32332.

Cohen AS, Stone JR, Beuning KRM *et al.* (2007) Ecological consequences of early Late Pleistocene megadroughts in tropical Africa. *Proceeding of the National Academy of Sciences, USA*, **104**, 16422–16427.

Colangelo P, Granjon L, Taylor PJ, Corti M (2007) Evolutionary systematics in African gerbilline rodents of the genus *Gerbilliscus*: inference from mitochondrial genes. *Molecular Phylogenetics and Evolution*, **42**, 797–806.

Colangelo P, Verheyen E, Leirs H *et al.* (2013) A mitochondrial phylogeographic scenario for the most widespread African rodent, *Mastomys natalensis*. *Biological Journal of the Linnean Society*, **108**, 901–916.

Cotterill FPD (2003) Geomorphological influences on vicariant evolution in some African mammals in the Zambezi Basin: some lessons for conservation. In: *Proceedings of the Ecology and Conservation of Mini-Antelope: An International Symposium on Duiker and Dwarf Antelope in Africa* (ed. Plowman A), pp. 11–58. Filander Verlag, Fürth.

Cotterill FPD (2006) Taxonomic status and conservation importance of the avifauna of Katanga (southeast Congo Basin) and its environs. *Ostrich*, **77**, 1–21.

Coulthard TJ, Ramirez JA, Barton N, Rogerson M, Brücher T (2013) Were rivers flowing across the Sahara during the Last Interglacial? Implications for human migration through Africa. *PLoS ONE*, **8**, e74834.

Darriba D, Taboada GL, Doallo R, Posada D (2012) JMODELTEST 2: more models, new heuristics and parallel computing. *Nature Methods*, **9**, 772.

Dempster ER (2013) *Gerbilliscus leucogaster*. In: *Mammals of Africa. vol. 3: Rodentia* (ed. Happold DCD), pp. 279–281. University of California Press, Berkeley, California.

Denys C (1999) Of mice and men. Evolution in East and South Africa during Plio-Pleistocene times. In: *African Biogeography, Climate Change & Human Evolution* (eds Bromage GT, Schrenk F), pp. 226–252. Oxford University Press, New York, New York.

Dobigny G, Tatard C, Gauthier P *et al.* (2013) Mitochondrial and nuclear genes-based phylogeography of *Arvicanthis niloticus* (Murinae) and Sub-Saharan open habitats Pleis-

tocene history. *PLoS ONE*, **8**, e77815. doi:10.1371/journal.pone.0077815.

Dormann CF, Elith J, Bacher S *et al.* (2013) Collinearity: a review of methods to deal with it and a simulation study evaluating their performance. *Ecography*, **36**, 27–46. doi:10.1111/j.1600-0587.2012.07348.x.

Drake NA, Blench RM, Armitage SJ, Bristow CS, White KH (2011) Ancient watercourses and biogeography of the Sahara explain the peopling of the desert. *Proceedings of the National Academy of Sciences, USA*, **108**, 458–462.

Drummond AJ, Suchard MA, Xie D, Rambaut A (2012) Bayesian phylogenetics with BEAUTi and the BEAST 1.7. *Molecular Biology and Evolution*, **29**, 1969–1973.

Earl DA, vonHoldt BM (2012) STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conservation Genetics Resources*, **4**, 359–361.

Edwards S, Claude J, Jansen Van Vuuren B, Matthee CA (2011) Evolutionary history of the Karoo bush rat, *Myotomys unisulcatus* (Rodentia, Muridae): disconcordance between morphology and genetics. *Biological Journal of the Linnean Society*, **102**, 510–526.

Escobar García P, Winkler M, Flatscher R *et al.* (2012) Extensive range persistence in peripheral and interior refugia characterizes Pleistocene range dynamics in a widespread Alpine plant species (*Senecio carniolicus*, Asteraceae). *Molecular Ecology*, **21**, 1255–1270.

Evanno G, Regnaut S, Goudet J (2005) Detecting the number of clusters of individuals using the software structure: a simulation study. *Molecular Ecology*, **14**, 2611–2620.

Excoffier L, Laval G, Schneider S (2005) ARLEQUIN ver. 3.0: an integrated software package for population genetics data analysis. *Evolutionary Bioinformatics Online*, **1**, 47–50.

Fahey AL, Ricklefs RE, Dewoody JA (2014) DNA-based approaches for evaluating historical demography in terrestrial vertebrates. *Biological Journal of the Linnean Society*, **112**, 367–386.

Faith JT, Behrensmeyer AK (2013) Climate change and faunal turnover: testing the mechanics of the turnover-pulse hypothesis with South African fossil data. *Paleobiology*, **39**, 609–627.

Falush D, Stevens M, Pritchard JK (2003) Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. *Genetics*, **164**, 1567–1587.

Falush D, Stevens M, Pritchard JK (2007) Inference of population structure using multilocus genotype data: dominant markers and null alleles. *Molecular Ecology Notes*, **7**, 574–578.

Faulkes CG, Bennett NC, Cotterill FPD, Stanley W, Mgode GF, Verheyen E (2011) Phylogeography and cryptic diversity of the solitary-dwelling silvery mole-rat, genus *Heliophobius* (family: Bathyergidae). *Journal of Zoology*, **285**, 324–338.

Feakins SJ, deMenocal PB, Eglinton TI (2005) Biomarker records of late Neogene changes in northeast African vegetation. *Geology*, **33**, 977–980.

Felsenstein J (1988) Phylogenies from molecular sequences: inference and reliability. *Annual Review of Genetics*, **22**, 521–565.

Fennessy J, Bock F, Tutchings A, Brenneman R, Janke A (2013) Mitochondrial DNA analyses show that Zambia's South Luangwa Valley giraffe (*Giraffa camelopardalis thornicrofti*) are genetically isolated. *African Journal of Ecology*, **51**, 635–640.

Fu YX (1997) Statistical tests of neutrality of mutations against population growth, hitchhicking and background selection. *Genetics*, **147**, 915–925.

Gascuel O (1997) BIONJ: an improved version of the NJ algorithm, based on a simple model of sequence data. *Molecular Biology and Evolution*, **14**, 685–695.

Gernhard T (2008) The conditioned reconstruction process. *Journal of Theoretical Biology*, **253**, 769–778.

Granjon L, Colangelo P, Tatard C, Colyn M, Dobigny G, Nicolas V (2012) Intrageneric relationships within *Gerbilliscus* (Rodentia, Muridae, Gerbillinae), with characterization of an additional West African species. *Zootaxa*, **3325**, 1–25.

Grubb P, Sandrock O, Kullmer O, Kaiser TM, Schrenk F (1999) Relationships between eastern and southern mammal faunas. In: *African Biogeography, Climate Change & Human Evolution* (eds Bromage GT, Schrenk F), pp. 253–267. Oxford University Press, New York, New York.

Guindon S, Dufayard JF, Lefort V, Anisimova M, Hordijk W, Gascuel O (2010) New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Systematic Biology*, **59**, 307–321.

Heled J, Drummond AJ (2008) Bayesian inference of population size history from multiple loci. *BMC Evolutionary Biology*, **8**, 289. doi:10.1186/1471-2148-8-289.

Hijmans RJ, Cameron SE, Parra JL, Jones PG, Jarvis A (2005) Very high resolution interpolated climate surfaces for global land areas. *International Journal of Climatology*, **25**, 1965–1978.

Jakobsson M, Rosenberg NA (2007) CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. *Bioinformatics*, **23**, 1801–1806.

Lemey P, Rambaut A, Welch JJ, Suchard MA (2010) Phylogeography takes a relaxed random walk in continuous space and time. *Molecular Biology and Evolution*, **27**, 1877–1885.

Librado P, Rozas J (2009) DNASP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics*, **25**, 1451–1452.

Lorenzen ED, Heller R, Siegismund HR (2012) Comparative phylogeography of African savannah ungulates. *Molecular Ecology*, **21**, 3656–3670.

Maslin MA, Pancost RD, Wilson KE, Lewis J, Trauth MH (2012) Three and half million year history of moisture availability of South West Africa: evidence from ODP site 1085 biomarker records. *Palaeogeography, Palaeoclimatology, Palaeoecology*, **317**, 41–47.

McDonough MM, Ammerman LK, Timm RM, Genoways HH, Larsen PA, Baker RJ (2008) Speciation within Bonneted bats (genus *Eumops):* the complexity of morphological, mitochondrial, and nuclear data sets in systematics. *Journal of Mammalogy*, **89**, 1306–1315.

McDonough M (2013) *Molecular systematics and phylogeography of the genus Gerbilliscus (Mammalia, Rodentia) in sub-Saharan Africa*. PhD Dissertation, Texas Tech University, Lubbock.

deMenocal PB (2004) African climate change and faunal evolution during the Pliocene-Pleistocene. *Earth and Planetary Science Letters*, **220**, 3–24.

Merow C, Smith MJ, Silander JA (2013) A practical guide to MaxEnt for modeling species' distributions: what it does, and why inputs and settings matter. *Ecography*, **36**, 1058–1069.

Mikula O, Šumbera R, Aghová T, Mbau J, Bryja J (in review) Evolutionary history and species diversity of African Pouched Mice of the genus *Saccostomus* (Rodentia: Nesomyidae). *Zoologica Scripta*.

Minnin VN, Bloomquist EW, Suchard MA (2008) Smooth skyride through a rough skyline: Bayesian coalescent-based inference of population dynamics. *Molecular Biology and Evolution*, **25**, 1459–1471.

Montoya M (2007) The last interglacial. In: *Encyclopedia of Quaternary Science* (ed. Elias SA), pp. 1916–1927. Elsevier, Amsterdam.

Moodley Y, Bruford MW (2007) Molecular Biogeography: towards an Integrated Framework for Conserving Pan-African Biodiversity. *PLoS ONE*, **2**, e454. doi:10.1371/journal.pone.0000454.

Moore AE, Cotterill FPD, Main MPL, Williams HB (2007) The Zambezi river. In: *Large Rivers: Geomorphology and Management* (ed. Gupta A), pp. 311–332. John Wiley & Sons, West Sussex, UK.

Moore AE, Cotterill FPD, Eckardt FD (2012) The evolution and ages of Makgadikgadi Palaeo-lakes: consilient evidence from Kalahari Drainage Evolution. *South African Journal of Geology*, **115**, 385–413.

Musser GG, Carleton MD (2005) Superfamily muroidea. In: *Mammal Species of the World: A Taxonomic and Geographic Reference*, 3rd edn (eds Wilson DE, Reeder DM), pp. 894–1531. Johns Hopkins University Press, Baltimore, Maryland.

Nascimento FF, Lazar A, Menezes AN *et al.* (2013) The role of historical barriers in the diversification processes in open vegetation formations during the Miocene/Pliocene using an ancient rodent lineage as a model. *PLoS ONE*, **8**, e61924.

Nei M (1972) Genetic distance between populations. *American Naturalist*, **106**, 283–292.

Otto-Bliesner BL, Marshall SJ, Overpeck JT *et al.* (2006) Simulating Arctic climate warmth and icefield retreat in the Last Interglaciation. *Science*, **311**, 1751–1753.

Owens HL, Campbell LP, Dornak LL *et al.* (2013) Constraints on interpretation of ecological niche models by limited environmental ranges on calibration areas. *Ecological Modelling*, **263**, 10–18.

Peakall R, Smouse PE (2006) GENALEX 6: genetic analysis in Excel. Population genetic structure software for teaching and research. *Molecular Ecology Notes*, **6**, 288–295.

Phillips SJ, Anderson RP, Schapire RE (2006) Maximum entropy modeling of species geographic distributions. *Ecological Modelling*, **190**, 231–259.

Potts R (2013) Hominin evolution in settings of strong environmental variability. *Quaternary Science Reviews*, **73**, 1–13.

Pritchard JK, Stephans M, Donnelly P (2000) Inference of population structure using multilocus genotype data. *Genetics*, **155**, 945–959.

Rambaut A (2009) *FigTree v1.3.1*. from http://tree.bio.ed.ac.uk/software/figtree/. Accessed 8 August 2010.

Rambaut A, Drummond AJ (2007) *TRACER v1.5*. Available from http://tree.bio.ed.ac.uk/software/tracer. Accessed 2 December 2009.

Ramírez-Soriano A, Ramos-Onsins SE, Rozas J, Calafell F, Navarro A (2008) Statistical analysis of neutrality tests under demographic expansions, contractions, and bottle- necks with recombination. *Genetics*, **179**, 555–567.

Rogers AR, Harpending H (1992) Population growth makes waves in the distribution of pairwise genetic differences. *Molecular Biology and Evolution*, **9**, 552–569.

Ronquist F, Huelsenbeck JP, Van der Mark P (2005) *MrBayes 3.1 Manual*. Available from http://mrbayes.csit.fsu.edu/ Accessed 17 July 2012.

Ronquist F, Teslenko M, van der Mark P *et al.* (2012) MrBayes 3.2: Efficient Bayesian Phylogenetic Inference and Model Choice Across a Large Model Space. *Systematic Biology*, **61**, 539–542.

Rosenberg NA (2004) Distruct: a program for the graphical display of population structure. *Molecular Ecology Notes*, **4**, 137–138.

Russo I-RM, Chimimba CT, Bloomer P (2010) Bioregion heterogeneity correlates with extensive mitochondrial DNA diversity in the Namaqua rock mouse, *Micaelamys namaquensis* (Rodentia: Muridae) from southern Africa – evidence for a species complex. *BMC Evolutionary Biology*, **10**, 307. doi:10.1186/1471-2148-10-307.

Schenekar T, Weiss S (2011) High rate of calculation errors in mismatch distribution analysis results in numbers false inferences of biological importance. *Heredity*, **107**, 511–512.

Scholz CA, Johnson TC, Cohen AS *et al.* (2007) East African megadroughts between 135 and 75 thousand years ago and bearing on early-modern human origins. *Proceedings of the National Academy of Sciences, USA*, **104**, 16416–16421.

Senut B, Pickford M, Mein P, Conroy G, Van Couvering J (1992) Discovery of 12 new Late Cainozoic fossiliferous sites in palaeokarsts of the Otavi Mountains, Namibia. *Comptes Rendus de L'Académie des Sciences*, **314**, 727–733.

Soberón J, Peterson AT (2005) Interpretation of models of fundamental ecological niches and species' distributional areas. *Biodiversity Informatics*, **2**, 1–10.

Soberón J (2007) Grinnellian and Eltonian niches and geographic distributions of species. *Ecology Letters*, **10**, 1115–1123.

Suchard MA, Weiss RE, Sinsheimer JS (2001) Bayesian selection of continuous-time Markov chain evolutionary models. *Molecular Biology and Evolution*, **18**, 1001–1013.

Tajima F (1989) Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics*, **123**, 585–595.

Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S (2011) MEGA5: molecular Evolutionary Genetics Analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Molecular Biology and Evolution*, **28**, 2731–2739.

Toews DP, Brelsford A (2012) The biogeography of mitochondrial and nuclear discordance in animals. *Molecular Ecology*, **21**, 3907–3930.

Trauth MH, Maslin MA, Deino AL *et al.* (2010) Human evolution in a variable environment: the amplifier lakes of Eastern Africa. *Quaternary Science Reviews*, **29**, 2981–2988.

Van Daele PAAG, Dammann P, Meier JL, Kawalika M, Van De Woestijne C, Burda H (2004) Chromosomal diversity in mole-rats of the genus *Cryptomys* (Rodentia: Bathyergidae) from the Zambezian region: with descriptions of new karyotypes. *Journal of Zoology, London*, **264**, 317–326.

Van Daele PAAG, Faulkes CG, Verheyen E, Adriaens D (2007) African mole-rats (Bathyergidae): a complex radiation in

tropical soils. In: *Subterranean Rodents: News From the Underground* (eds Begall S, Burda H, Schleich CE), pp. 357–373. Springer, Berlin.

Varela S, Rodríguez S, Lobo JM (2009) Is current climatic equilibrium a guarantee for the transferability of distribution model predictions? A case study of the spotted hyena. *Journal of Biogeography*, **36**, 1645–1655.

Werdelin L, Sanders WJ (2010) *Cenozoic Mammals of Africa*. University of Californian Press, Berkley, California.

Wesselman HB, Black MT, Asnake M (2009) Small mammals. In: *Ardipithecus Kadabba: Late Miocene Evidence from the Middle Awash, Ethiopia (Vol. 2)* (eds Selassie YH, WoldeGabriel G), pp. 105–133. University of California Press, Berkeley, California.

White RM, Kerley GIH, Bernard RTF (1997) Pattern and controls of reproduction of the southern African rodent Gerbillurus paebain the semi-arid Karoo, South Africa. *Journal of Arid Environments*, **37**, 529–549.

Woltering M, Johnson TC, Werne JP, Schouten S, Sinninghe Damsté JS (2011) Late Pleistocene temperature history of Southeast Africa: a TEX_{86} temperature record from Lake Malawi. *Palaeogeography, Palaeoclimatology, Palaeoecology*, **303**, 93–102.

Zinner D, Groeneveld LF, Keller C, Roos K (2009) Mitochondrial phylogeography of baboons (*Papio* spp.) – Indication for introgressive hybridization? *BMC Evolutionary Biology*, **9**, 83. doi:10.1186/1471-2148-9-83.

J.B., M.M.M. and R.S. conceived and funded the project. J.B. and M.M.M. conducted the genetic laboratory work and data analyses. A.W.F. and V.M. performed E.N.M. analyses. M.M.M., R.S., A.W.F., V.M. and J.B. contributed to fieldwork. All authors contributed to writing of the manuscript.

Data accessibility

DNA sequences are available under the GenBank Accession nos KM453986-KM454070 (Cytb), KM454071-KM454112 (TKTL1) and KM454113-KM454140 (UTY11). Sampling locations and additional individuals included in ENMs are uploaded as online Supporting Information. Sequence alignments, AFLP genotype data and parameter files archived on Dryad (http://datadryad.org), doi:10.5061/dryad.83kd9.

Supporting information

Additional supporting information may be found in the online version of this article.

Fig. S1 Depiction of boundaries for four training areas used in ENMs.

Fig. S2 Plot for principal coordinate analysis depicting PC2 and PC3.

Fig. S3 Median joining networks for X- and Y- chromosome markers.

Fig. S4 Maxent models using different training areas.

Fig. S5 Comparative ecological niche models for the present, LGM and LIG projected to continental Africa.

Table S1 Parameter input and output for calculating demographic expansion.

Appendix S1 Individuals included in genetic analyses.

Appendix S2 Molecular methods.

Appendix S3 Individuals included in the ENM analysis.

Chapter III

Pan-African phylogeny of *Mus* (subgenus *Nannomys*) reveals one of the most successful mammal radiations in Africa.

Bryja J., Mikula O., Šumbera R., Meheretu Y., Aghová T., Lavrenchenko L. A., **Mazoch V.**, Oguge N., Mbau J. S., Welegerima K., Amundala N., Colyn M., Leirs H. & Verheyen E. (2014). *BMC Evolutionary Biology*, 14(1), 256.

Bryja et al. BMC Evolutionary Biology (2014) 14:256
DOI 10.1186/s12862-014-0256-2

BMC Evolutionary Biology

RESEARCH ARTICLE Open Access

Pan-African phylogeny of *Mus* (subgenus *Nannomys*) reveals one of the most successful mammal radiations in Africa

Josef Bryja[1,2,13*], Ondřej Mikula[1,3], Radim Šumbera[4], Yonas Meheretu[5], Tatiana Aghová[1,2], Leonid A Lavrenchenko[6], Vladimír Mazoch[4], Nicholas Oguge[7], Judith S Mbau[8], Kiros Welegerima[5], Nicaise Amundala[9], Marc Colyn[10], Herwig Leirs[11] and Erik Verheyen[11,12]

Abstract

Background: Rodents of the genus *Mus* represent one of the most valuable biological models for biomedical and evolutionary research. Out of the four currently recognized subgenera, *Nannomys* (African pygmy mice, including the smallest rodents in the world) comprises the only original African lineage. Species of this subgenus became important models for the study of sex determination in mammals and they are also hosts of potentially dangerous pathogens. *Nannomys* ancestors colonized Africa from Asia at the end of Miocene and Eastern Africa should be considered as the place of their first radiation. In sharp contrast with this fact and despite the biological importance of *Nannomys*, the specimens from Eastern Africa were obviously under-represented in previous studies and the phylogenetic and distributional patterns were thus incomplete.

Results: We performed comprehensive genetic analysis of 657 individuals of *Nannomys* collected at approximately 300 localities across the whole sub-Saharan Africa. Phylogenetic reconstructions based on mitochondrial (*CYTB*) and nuclear (*IRBP*) genes identified five species groups and three monotypic ancestral lineages. We provide evidence for important cryptic diversity and we defined and mapped the distribution of 27 molecular operational taxonomic units (MOTUs) that may correspond to presumable species. Biogeographical reconstructions based on data spanning all of Africa modified the previous evolutionary scenarios. First divergences occurred in Eastern African mountains soon after the colonization of the continent and the remnants of these old divergences still occur there, represented by long basal branches of *M.* (previously *Muriculus*) *imberbis* and two undescribed species from Ethiopia and Malawi. The radiation in drier lowland habitats associated with the decrease of body size is much younger, occurred mainly in a single lineage (called the minutoides group, and especially within the species *M. minutoides*), and was probably linked to aridification and climatic fluctuations in middle Pliocene/Pleistocene.

Conclusions: We discovered very high cryptic diversity in African pygmy mice making the genus *Mus* one of the richest genera of African mammals. Our taxon sampling allowed reliable phylogenetic and biogeographic reconstructions that (together with detailed distributional data of individual MOTUs) provide a solid basis for further evolutionary, ecological and epidemiological studies of this important group of rodents.

Keywords: Biogeography, Tropical Africa, Molecular phylogeny, Pygmy mice, Plio-Pleistocene climatic fluctuations, Divergence timing, Muridae (Murinae), *Mus minutoides*, Phylogeography, DNA barcoding

* Correspondence: bryja@brno.cas.cz
[1]Institute of Vertebrate Biology, Academy of Sciences of the Czech Republic, Brno, Czech Republic
[2]Department of Botany and Zoology, Faculty of Science, Masaryk University, Brno, Czech Republic
Full list of author information is available at the end of the article

Background

One of the main challenges of current nature conservation is the accelerating loss of biodiversity. Even if this problem is generally recognized, there are several difficulties in quantifying the loss of biodiversity at the level of species. For example, there is a lack of traditional taxonomic specialists for particular groups of organisms and the real amount of biodiversity is therefore unknown [1]. This is especially true for some tropical areas, where the overall biodiversity level is the highest and its loss is the most intensive. Another problem for practical biodiversity conservation is the delimitation of species (e.g. [2] vs. [3]). Traditional concepts of typological or biological species are not universally applicable and with accumulating knowledge in evolutionary biology it is increasingly difficult to define generally what a species is. Genetic approaches, like DNA barcoding, are now routinely used to overcome some of these problems. They provide a cheap and easily applicable approach for discovering the taxa worth future taxonomical research and areas with high phylogenetic diversity with special conservation concern (e.g. [4]). For example, 175 new extant taxa of mammals were described from African mainland, Madagascar and all surrounding islands between 1988–2008 [5], and in the majority, the first consideration for taxonomic delimitation was motivated by the use of genetic data.

Rodents of the genus *Mus* represent one of the most valuable biological models for biomedical and evolutionary research [6]. Out of the four currently recognized subgenera, i.e. *Mus*, *Coelomys*, *Pyromys* and *Nannomys*, the latter comprises the African pygmy mice [7]. These are small rodents (4–12 g in most taxa, but see [8]), endemic to the sub-Saharan Africa. The phylogenetic relationships, species diversity, ecology and chromosomal evolution of *Nannomys* were recently reviewed [9]. They represent the most diverse lineage of the genus, with currently about 18 species recognized [9,10], comprising almost half of the described *Mus* species [10]. While predominantly savannah dwellers [11], several species have also been trapped in forest, agricultural fields and rural areas [12-14].

Mainly due to their extensive chromosomal diversity coupled with highly conserved morphology, African pygmy mice have attracted the attention of evolutionary scientists [9,11,15-17]. They exhibit chromosomal features that are rarely recorded in other taxa, e.g. the greatest diversity of sex-autosome translocations reported so far in any mammalian lineage (e.g. [18]). Thus *Nannomys* became an important biological model for the study of processes of chromosomal speciation and mechanisms of sex determination in mammals [19]. Recent studies have also shown that African pygmy mice are important hosts of arena viruses [20-23], making them a target group for epidemiological surveys.

Increasing numbers of molecular genetic data provide evidence for high cryptic diversity in *Nannomys* and it is highly probable that further integrative taxonomy research will reveal new undescribed species [11,14]. Furthermore, the inclusion of poorly known African *Mus*-related rodents in molecular phylogenetic datasets may provide surprising results changing the current view on the evolutionary scenarios of *Nannomys*. For example, the Ethiopian endemic genus *Muriculus* was recently recognized to be an internal lineage of *Mus* [8].

The genus *Mus* diverged in Asia approximately 6.7 to 7.8 Mya and shortly after this time the ancestor(s) of *Nannomys* colonized Africa through the Arabian Peninsula and Miocene land bridges [9]. The oldest fossils of *Mus* in Africa are reported from Tugen Hills (Kenya) about 4.5 Mya [24]. The highly heterogeneous environment of Eastern Africa can thus be considered as the place of first diversification of African *Mus* in Early Pliocene, followed by a radiation caused by climatic oscillations and habitat modification [9,11]. In this context it is important to note that genetic data used so far for molecular phylogenetic inference of the African pygmy mice are strongly biased geographically in favour of material collected from savannahs in the western and southern part of the continent, while specimens from Eastern Africa (including those from mountains and forests) are clearly under-represented (Figure 1a).

More thorough geographical sampling is necessary for obtaining the correct biogeographical scenario of *Nannomys* evolution. Only a comprehensive and reliable phylogenetic hypothesis can lead to meaningful inferences on the evolution of sex-determination or virus-host co-evolution. In this study, we provide the so far most comprehensive geographic sampling of genetically characterized African pygmy mice composed of 657 *Nannomys* individuals from most parts of sub-Saharan Africa. First, we use this pan-African dataset for the reconstruction of phylogenetic relationships within *Nannomys* lineage. Second, using the combination of species delimitation methods, we aim to estimate the presumable species richness of *Nannomys*, highlighting groups and geographical regions necessitating further taxonomical research. Finally, the dating of divergences and biogeographical reconstructions allow us to modify previous scenarios that were suggested to explain the *Nannomys* radiation in Africa.

Methods

Sampling

New genetic data were produced for 395 individuals of subgenus *Nannomys* sampled in sub-Saharan Africa by the authors and their collaborators. All fieldwork complied with legal regulations in particular African countries and sampling was in accordance with local legislation (see more details on wildlife authorities that permitted the

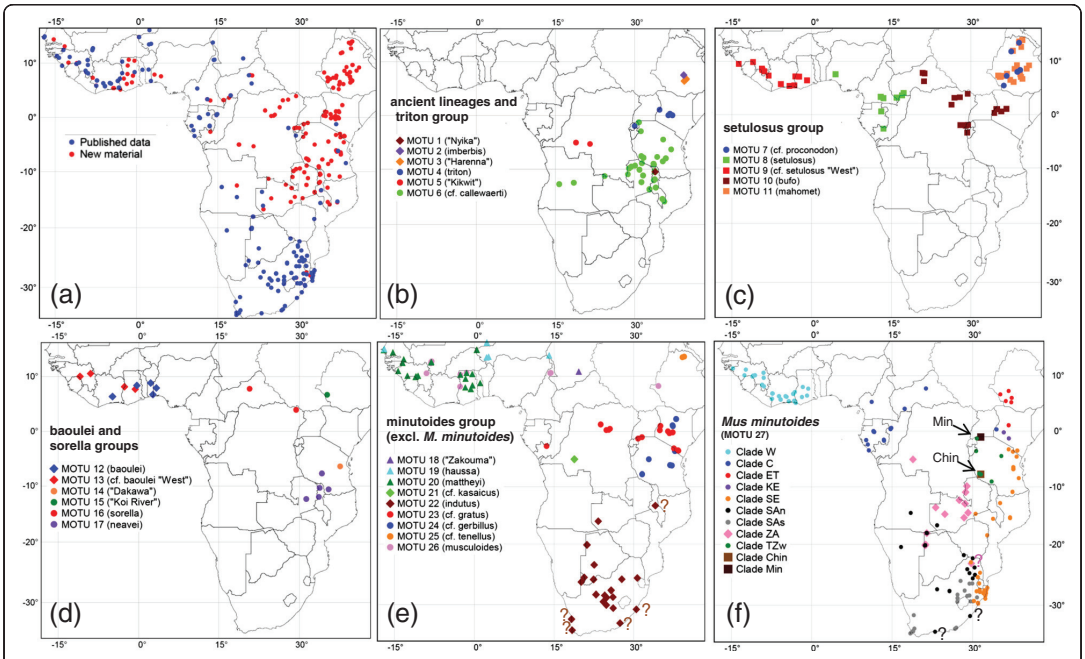


Figure 1 Distribution of genotyped specimens and individual MOTUs. (a) Distribution of analyzed material of *Nannomys* in Africa. Blue dots indicate the geographical position of published sequences (downloaded from GenBank), red dots show the localities of newly sequenced individuals. Geographical distribution of **(b)** MOTUs of the triton group (circles) and three ancient monotypic lineages (rhombuses); **(c)** MOTUs of the setulosus group; **(d)** MOTUs of the baoulei (rhombuses) and the sorella (circles) groups; **(e)** MOTUs of the minutoides group (except *M. minutoides*); **(f)** phylogeographic structure of MOTU 27, i.e. *M. minutoides*. In Figures 1e identical symbol shapes represent monophyletic groups. In Figure 1f the clade abbreviations correspond to Figure 4. Question marks indicate doubtful records based on genotyping of old museum material (see [64]). For more information on analysed material see Additional file 1.

research in Acknowledgements). Each individual was identified to the genus by the external features and the tissue sample (tail, toe, spleen, etc.) was stored in 96% ethanol until DNA extraction. GPS coordinates of each locality were recorded. New data were supplemented with 262 published records of genotyped and georeferenced *Nannomys*, i.e. partial or complete sequences of mitochondrial gene for cytochrome *b* (*CYTB*) were downloaded from GenBank. Geographical coordinates of published data were either retracted from original publications (if available) or approximately estimated from Google maps. In total, the analysed dataset includes genetic information of 657 individuals from approximately 300 localities in 30 African countries (Figure 1a). For more details on analysed individuals see Additional file 1.

DNA sequencing of *CYTB* and *IRBP*

DNA was extracted using the DNeasy Blood & Tissue Kit (Qiagen). The complete *CYTB* gene was amplified by polymerase chain reaction (PCR) using primers L14723 and H15915 [25]. PCR mix contained 3 μl of genomic DNA, 0.5 units of Taq polymerase (Fermentas), final concentrations of 3 mM $MgCl_2$, 1 x Taq buffer with $(NH_4)_2SO_4$ (Fermentas), 0.2 mM of each dNTPs, 0.2 μM of each primers and ddH_2O to final volume of 30 μl. The thermal profile of the PCR started with an initial denaturation at 94°C for 3 min, followed by 35 cycles composed of 30 s of denaturation at 94°C, 30 s of annealing at 50°C, and 3 min of extension at 72°C and PCR was finished by a final extension at 72°C for 10 min. The part of nuclear gene encoding the Interphotoreceptor Binding Protein (*IRBP*) was amplified in selected individuals (from each main clade identified previously by *CYTB* marker) by the primers IRBP1531 and IRBP217 [26]. PCR conditions were the same as above, except the final concentration of $MgCl_2$ (2 mM). The thermal profile of the PCR started with an initial denaturation of one step at 94°C for 3 min, followed by 30 cycles of 60 s at 94°C, 60 s at 55°C, 2 min at 72°C and finished by a final extension at 72°C for 10 min. The PCR products were purified with Calf Intestine Alkaline Phosphatase (ThermoScientific) and Exonuclease I (ThermoScientific) and sequenced along both strands commercially in Macrogen Europe using the same primers as for the PCR. Both genetic

markers have been previously used to successfully resolve systematic relationships in a wide range of related murid rodents (e.g. [7,25].

Genetic data from fresh material were complemented by museum samples (mostly dry skins) from the Royal Museum for Central Africa (Tervuren, Belgium), Muséum National d'Histoire Naturelle (Paris, France), American Museum of Natural History (New York, USA) and Hungarian Natural History Museum (Budapest, Hungary) (for more details see Additional file 1). Museum samples comprised only minor part of analysed material and we used especially those from geographical areas difficult to be accessed today (e.g. Central African Republic, eastern Democratic Republic of Congo) and the type material of *Mus bufo*. All museum samples were handled in a specialized laboratory of Institute of Vertebrate Biology ASCR in Studenec, designed for work with rare DNA to prevent contamination by samples with high quantity of DNA or PCR products. DNA was extracted using the JETQUICK Tissue DNA Spin Kit (Genomed). PCR amplification and pyrosequencing on GS Junior were performed according to mini-barcode protocol described by [27]. The main advantage of this approach in analysis of museum samples is that it allows for separating individual sequences in samples contaminated by distantly related organisms (e.g. contamination by human DNA), which is not possible through the Sanger sequencing.

Phylogenetic reconstructions

Sequences of *CYTB* and *IRBP* were edited and aligned in SeqScape v2.5 (Applied Biosystems), producing a final alignment of 1140 and 1276 bp, respectively. The Findmodel web application (http://www.hiv.lanl.gov/content/sequence/findmodel/findmodel.html) was used to identify the most appropriate substitution model for each gene. The Akaike information criterion (AIC), compared among 12 substitution models, revealed that the model best fitting the ingroup data was the General time reversible model with a gamma-distributed rate variation across sites (GTR + G) for both *CYTB* and *IRBP*. As outgroups, we used sequences of four species from other subgenera of the genus *Mus*, i.e. *M. platythrix* (*CYTB* GenBank Acc. code AJ698880, *IRBP* GenBank Acc. code AJ698895), *M. pahari* (AY057814, AJ698893), *M. caroli* (AB033698, AJ698885) and *M. musculus* (V00711, AF126968); two sister lineages of the genus *Mus* within subfamily Murinae, i.e. *Apodemus flavicollis* (AB032853, AB032860) and *Ratus norvegicus* (V01556, AJ429134); and one species from the subfamily Acomyinae, *Acomys cahirinus* (AJ233953, AJ698898) see also [7,9,11].

Phylogenetic relationships within *Nannomys* were inferred by maximum likelihood (ML) and Bayesian (BI) approaches. ML analysis was performed using RAxML 8.0 [28]. The GTR + G model (option -m GTRGAMMA) was selected for the six partitions, i.e. 1140 bp of *CYTB*, 1276 bp of *IRBP*, and both genes were partitioned also by the position of nucleotides in the codons (option -q). The robustness of the nodes was evaluated by the default bootstrap procedure with 1,000 replications (option -# 1000). Bayesian analysis of evolutionary relationships was performed by Markov chain Monte Carlo (MCMC) method in MrBayes v. 3.2.1 [29]. Three heated and one cold chain were employed in all analyses, and runs were initiated from random trees. Two independent runs were conducted with 5 million generations per run; and trees and parameters were sampled every 1,000 generations. Convergence was checked using TRACER v1.5 [30]. For each run, the first 10% of sampled trees were discarded as burn-in. Bayesian posterior probabilities were used to assess branch support of the Bayesian tree.

The most widespread *Nannomys* species (= MOTU, see below) is *M. minutoides*. For this species we performed more detailed analysis of intraspecific genetic variability. We selected 131 sequences belonging to this clade and trimmed the final alignment to the length of 741 bp. Haplotypes were generated using DNaSP software [31] and a median-joining network of haplotypes was produced in the software Network 4.6.1.2 (downloaded on 10.2.2014 from http://www.fluxus-engineering.com/sharenet.htm).

Delimitation of MOTUs

We estimated the possible number of putative species (called here molecular operational taxonomic units, MOTUs, until the thorough taxonomic evidence will be provided) of *Nannomys* in the sampled dataset by using two types of divergence thresholds and the *CYTB* dataset. The first was the time threshold estimated by the Generalized Mixed Yule Coalescent (GMYC) model [32] which describes single-locus branching pattern as a succession of speciation events replaced at a fixed threshold time by a succession of intraspecific coalescent events. The two stages are modelled by Yule process and neutral coalescent, respectively, which allows finding maximum likelihood estimate of the threshold time and evaluating statistical support for the delimited species [33,34]. In this framework reliably delimited species are those whose basal internal split occurred well after the speciation-coalescence threshold and which diverged from sister species well before it. We therefore calculated two kinds of support: (1) for each intra-specific basal split we calculated relative likelihood that it represents coalescence rather than speciation event by summing up Akaike weights of all threshold times older or equal to its age; (2) for each inter-specific split we calculated relative likelihood that it represents speciation as a sum of Akaike weights of threshold times younger to it. The ultrametric tree required by GMYC was produced by BEAST 1.8.0 [35] with uncorrelated lognormal distribution of substitution

rates and lognormal priors for node ages mimicking posteriors from the divergence dating (see below). We used the Yule prior assuming no intra-specific divergences (alternative analyses with a coalescent prior assuming no speciation events lead to almost identical results of GMYC analyses; not shown). The topology was constrained to match the branching order of main lineages observed in the maximum likelihood phylogeny. The GMYC analysis was performed using the R package 'splits' (http://r-forge.r-project.org/R/?group_id=333).

The second threshold was based on sequence divergence, taken as a proxy for the amount of genetic difference among distinct gene pools. We therefore analyzed the distribution of Kimura-2 parameter (K2P) corrected genetic distances on *CYTB* among GMYC-delimited species (calculated in Mega 5.05; [36]) and merged the lineages with less than 7.3% genetic distance, i.e. the mean value between sister species of rodents [37]. The resulting groups were designated as molecular operational taxonomic units (MOTUs) and provisional names were assigned to them. It is important to note that the aim of our MOTUs delimitation approach is not to change the current taxonomy, but to highlight the taxa and geographical areas worthy of further taxonomic study, including morphological, ecological and more detailed genetic approaches.

Divergence dating

Time to the most recent common ancestors (TMRCA) of clades identified by phylogenetic analyses was estimated using a relaxed clock model with substitution rates drawn from an uncorrelated lognormal distribution in BEAST 1.8.0 [35] and three fossil-based calibration points: origin of extant *Mus*, origin of extant *Apodemus* and the *Arvicanthis*/*Otomys* lineage split. To avoid disproportionate impact of *Nannomys* we fitted the evolutionary model to 63 concatenated *CYTB* and *IRBP* sequences representing main lineages of *Nannomys* and correspondingly deep divergences across the tribes Apodemini, Arvicanthini, Malacomyini, Murini, Otomyini and Praomyini (sensu [38]). The data set is reported in detail in the Additional file 2.

Following [39] we used lognormal calibration densities with zero means whose 5% and 95% quantiles were specified by appropriately chosen standard deviations and offsets and corresponded to the fossil derived minimum and maximum ages. In particular the parameters (standard deviation, offset, 5% and 95% quantile) were: (1) 0.74, 7.00, 7.30 and 10.38 for *Mus*, based on the earliest fossil *Mus* and a member of *Progonomys* considered belonging to *Mus* stem lineage [40]; (2) 0.54, 4.89, 5.30 and 7.30 for *Apodemus* corresponding to 95% confidence interval of first appearance as reported by [39], although we applied it to the basal split of extant species rather than to the origin of their stem lineage; (3) 0.80, 5.81, 6.08 and 9.54 for *Arvicanthis*/*Otomys* split which was derived from the earliest records of *Otomys* (ca. 5 Mya; [41]) and arvicanthine genera *Aethomys*, *Arvicanthis* and *Lemniscomys* (6.08–6.12 Mya; [42]) and the next relevant sample where these and related genera (except for a tentative *Aethomys*) are absent (9.50-10.50 Mya; [43]). Based on the previous studies [38,39,44] we constrained the topology to include a basal split between Arvicanthini+Otomyini and the rest of the species.

The MCMC simulations were run twice with 25 million iterations, with genealogies and model parameters sampled every 1000 iterations. Trees were linked, models and clocks were unlinked for two markers. Convergence was checked using TRACER v1.5 [30], both runs being combined in LOGCOMBINER 1.7.1 [35] and the maximum clade credibility tree calculated by TREEANNOTATOR 1.7.1 [35], following the removal of 10% burn-in.

Biogeographical analysis

Ancestral habitat types were inferred by the Bayesian analysis of discrete traits [45]. It models discrete states of a trait at the end of each branch as a result of a continuous time Markov chain with infinitesimal transition rates determined by an overall transition rate, pair-wise transition probabilities and a base frequency of the states. Following the current implementation in BEAST 1.8.0 we used strict clock time-irreversible model so the overall transition rate was assumed uniform across the tree and transition probabilities were allowed to differ in the opposite directions. Using the distribution data, we coded the 27 MOTUs as inhabiting either (i) tropical forests in the Congo Basin, Central and Western Africa; (ii) mountains in Eastern Africa (various habitats), or (iii) savannah habitats in sub-Saharan Africa. Some MOTUs can inhabit more habitat types (e.g. MOTU 27, *M. minutoides*). The analysis in BEAST does not allow more variants of the tip trait, so we assigned the trait (habitat) that is the most widespread in a particular MOTU (e.g. savannah in *M. minutoides*). The topology was fixed to match relationships between MOTUs on the ML tree and branch lengths were time-calibrated as in the ultrametric tree for GMYC.

Alternatively, we identified ancestral habitat types and rough geographic ranges by using the maximum likelihood approach implemented in the software Lagrange [46,47]. The implemented model estimates geographic range evolution using a phylogenetic tree with branch lengths scaled to time, geographic (habitat) areas for all tips, and an adjacency matrix of plausibly connected areas. We used the same tree and distribution data as in the BEAST analysis described above. We allowed the connection between all three habitats with equal probabilities of each transition. The maximum number of

ancestral ranges was set to two. The resulting reconstructions returned all models within two likelihood units of the best model, which we summarized for each daughter branch and plotted in the form of pie-charts along the tree in R [48].

Results

Overview of collected data

For the phylogenetic analysis we retained 179 *CYTB* sequences at least 700 bp (133 new sequences and 46 sequences from GenBank) representing as complete a geographical distribution of each clade as possible (Additional file 1). The remaining 478 sequences (usually shorter and/or from the same or close neighbouring localities), including 16 sequences obtained by 454 pyrosequencing of old museum samples, were unambiguously assigned to particular MOTU by neighbour-joining analysis in MEGA 5.05 (bootstrap values higher than 95%) and these data were used for mapping the geographical distribution of phylogenetic clades.

We also selected 1–2 individuals from each of the main significantly supported *CYTB* clades (if the tissues were available) and sequenced them at *IRBP* gene. The final phylogenetic analyses included 42 sequences of *IRBP* (32 new sequences and 10 sequences from GenBank; see Additional file 1) from all main species groups except the baoulei group (see below). ML analyses were performed separately for both genes, and because the topology of trees was very similar (although the resolution of *IRBP* was much lower; Additional file 3), we finally performed both ML and BI reconstructions only using a concatenated *CYTB* and *IRBP* dataset produced in SEAVIEW [49].

Phylogeny of African *Nannomys*

Phylogenetic trees based on the concatenated dataset were well resolved and with very similar topology of 179 ingroup sequences in both ML and BI analyses (Figure 2). Subgenus *Nannomys* (including "*Muriculus*" *imberbis*; see [8]) was strongly supported. There are three long branches representing three ancient mountainous species with unresolved relationships to other groups (*M.* sp. "Nyika" = MOTU 1, *M. imberbis* = MOTU 2, and *M.* sp. "Harenna" = MOTU 3) and five well supported species groups. We call them hereafter triton, setulosus, baoulei, sorella, and minutoides groups, based on the previous use of these names, representing the best known species within particular clades. Each group contains several distinct lineages that may represent separate species; the most diversified is the minutoides group. The relationships among species groups are not well resolved, but in most topologies the triton group is non-significantly clustered with three ancient species, while all other species groups cluster together. Within the latter, the setulosus group separates the first, and the baoulei group is the sister of the sorella group (Figure 2).

Number of potential species and their distribution

The application of the GMYC model provided the delimitation of 49 maximum likelihood entities (hereafter GMYC-species; 95% CI = 42–62 entities) based on the ML estimate of speciation-coalescence threshold at 0.46 (0.27–0.86) Mya. Figure 3a depicts support for the "intraspecific" basal splits as coalescences as well as support for "interspecific" splits as speciation events. In both cases white circles indicate support < 0.95 and black circles > 0.95. Low "intraspecific" support suggests there may be more species present, whereas low "interspecific" support suggests the two sister clades may be in fact conspecific populations. Where two neighbouring "interspecific" and "intraspecific" supports are low, the speciation-coalescence transition is blurred.

K2P distances among the GMYC-species (3.16-20.77%) were not overlapping with "intraspecific" distances (0.12-2,38%) (Additional file 4). The detailed analysis of geographical distribution of GMYC-species showed that many sister groups among them are parapatric, i.e. most probably representing the results of allopatric differentiation and secondary contacts. For example, in the clade corresponding to *M. minutoides* in previous studies (e.g. [9]), the GMYC method delimited 12 GMYC-species with prevailing parapatric distribution pattern and with "interspecific" K2P distances 3.27-6.96%. Using the threshold value of 7.3%, we grouped these lineages and considered them as phylogeographical differentiation within the single species *M. minutoides* (see Figure 1f for the distribution of phylogeographical lineages that roughly correspond to "species" identified by GMYC method). Using this combined approach (i.e. analysis of geographic distribution of GMYC-species and threshold of K2P distances), we reduced 49 GMYC-species to 27 highly supported molecular operational taxonomic units (MOTUs, Figure 3a), which are further discussed below. Genetic distances among 27 MOTUs were always significantly higher and did not overlap with those within MOTUs (Additional file 4).

There were 17 MOTUs that exactly matched a single GMYC-species, 11 of them represented by more than one sequence. 7 MOTUs comprised two GMYC-species, 2 MOTUs were composed of three GMYC-species and a single MOTU, MOTU 27 = *M. minutoides*, comprised 12 GMYC-species (Figure 3a). In 12 cases, however, there was strong support for the presence of multiple species within a single MOTU (marked by black circles left of the GMYC threshold in Figure 3a).

Below we follow the nomenclature of [10] that recognizes 18 valid species. Possible names for newly recognized MOTUs are discussed in the text.

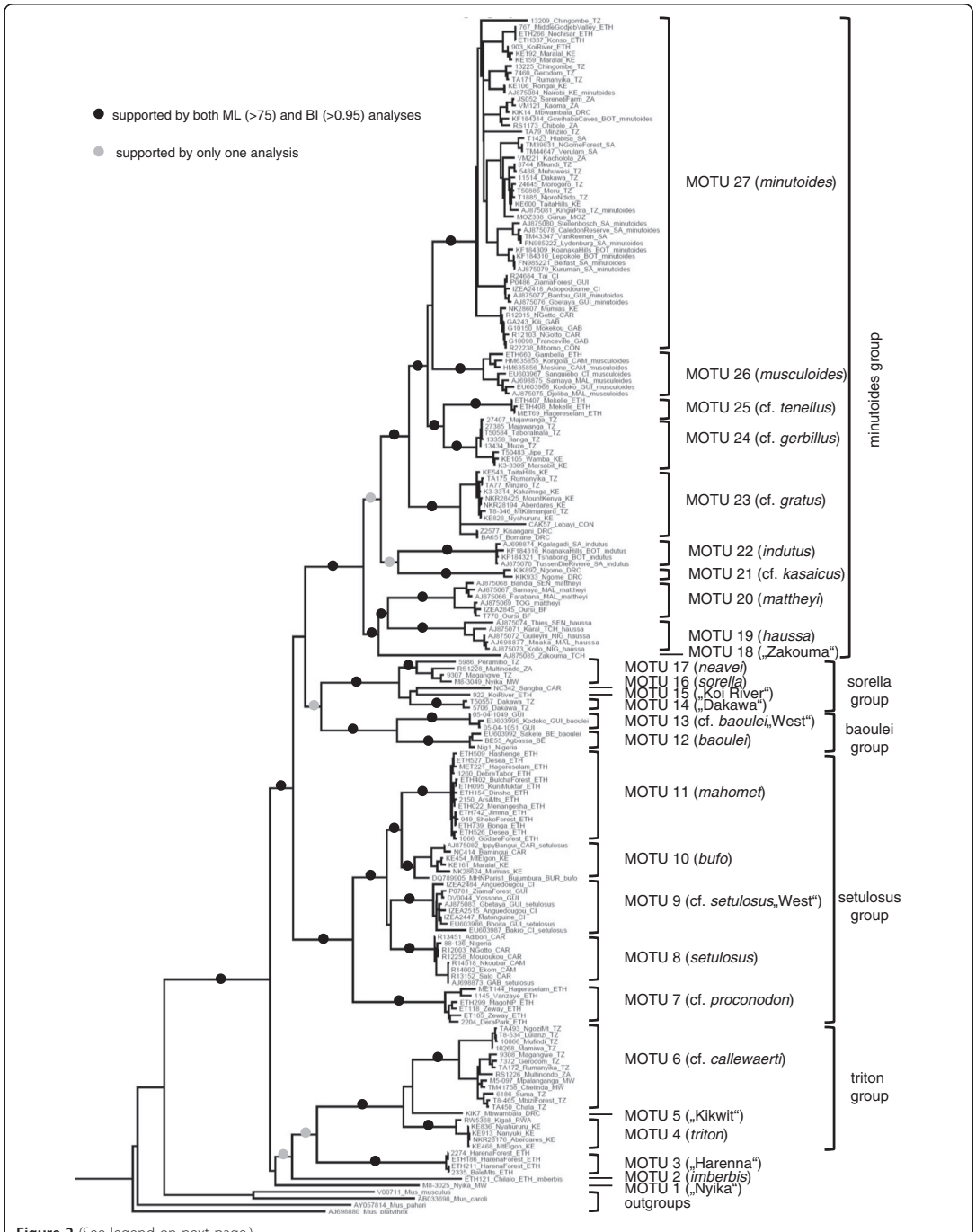


Figure 2 (See legend on next page.)

(See figure on previous page.)
Figure 2 Inferred phylogenetic relationships within *Nannomys*. Maximum likelihood phylogenetic tree of *Nannomys* is based on the combined dataset of mitochondrial (*CYTB*) and nuclear (*IRBP*) genes. Black circles indicate the support by both ML (bootstrap values > 75%) and BI (posterior probabilities > 0.95) analyses; grey circles indicate nodes supported by only one analysis. MOTUs were identified by the combination of GMYC approach and distribution of genetic distances on *CYTB*. Only outgroups from the genus *Mus* are shown. GenBank accession numbers correspond to *CYTB* sequences, for *IRBP* numbers see Additional file 1. Abbreviations of countries: BE: Benin, BF: Burkina Faso, BOT: Botswana, BUR: Burundi, CAM: Cameroon, CAR: Central African Republic, CI: Côte d'Ivoire, CON: Congo, DRC: Democratic Republic of Congo, ETH: Ethiopia, GAB: Gabon, GUI: Guinea, KE: Kenya, MAL: Mali, MOZ: Mozambique, MW: Malawi, NIG: Niger, RWA: Rwanda, SA: South Africa, SEN: Senegal, TOG: Togo, TCH: Tchad, TZ: Tanzania, ZA: Zambia.

(1)Ancient mountain lineages (Figure 1b):

A tri-phyletic group with very restricted distribution ranges. They are known from only a few individuals captured in the highest East African mountains. They were not included in previous phylogenetic studies of *Nannomys* and on the phylogenetic tree they form very long branches, in most topologies they are related to the triton group, but not always with significant nodal support.

(MOTU 1) *Mus* sp. "Nyika"

It is a very distinct ancient lineage of *Nannomys*, known from a single, relatively large individual (14 g), captured in the high plateau of Nyika Mts. in Malawi (cca 2100 m a.s.l.). Albeit partially broken, the cranium of this specimen clearly shows features that are typical for insectivorous rodents, namely proodont (forward oriented) incisors and slender mandibles. This lineage is sympatric with MOTU 17 (*M. neavei*) and even syntopic with MOTU 6 (*M.* cf. *callewaerti*).

(MOTU 2) *Mus imberbis* Rüppell, 1842

It is an easily distinguished taxon, very large (sequenced individual weighted 25 g) and with a black dorsal stripe. It has been considered as a separate genus *Muriculus*, but genetic analysis of a recently captured individual clearly shows that it is an internal lineage of *Mus* [8]. It is an endemic species of the high plateaux of Ethiopia, known only from a few of individuals (reviewed in [8]).

(MOTU 3) *Mus* sp. "Harenna"

It is a large species (cca 16 g), very probably endemic to the moist Harenna forest in the Bale Mts. in Ethiopia,

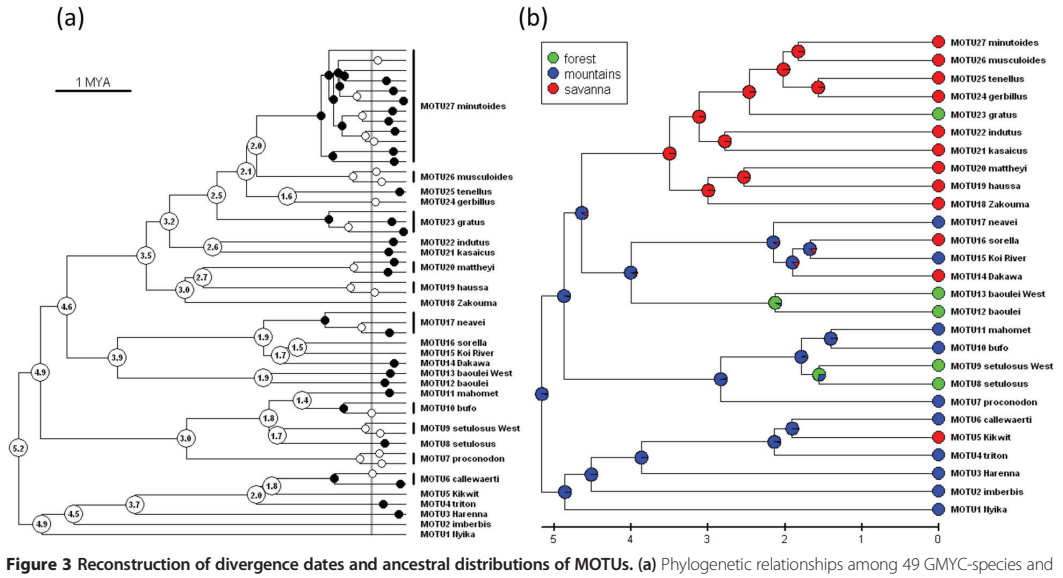


Figure 3 Reconstruction of divergence dates and ancestral distributions of MOTUs. (a) Phylogenetic relationships among 49 GMYC-species and definition of 27 MOTUs. The vertical line indicates the threshold where the speciation processes are replaced by coalescence. Black circles indicate strong support (>95%) for either speciation (left of the threshold) or intraspecific coalescence (right of the threshold). White circles indicate weak support (<95%) for these processes. The dating of divergences within *Nannomys* was assessed by BEAST using the previously estimated divergence times (see Additional file 2) as priors for calibration of relaxed molecular clock. **(b)** Reconstruction of ancestral distribution areas (blue – mountains in Eastern Africa, green – tropical forests of central and western Africa; red – open savannah-like habitats surrounding forests and mountains in sub-Saharan Africa. The different colours on pie charts indicate the probability of a particular state of the trait for each node. The analysis of ancestral traits was performed in BEAST (see text for more details).

a region with very pronounced endemicity [50,51]. Based on morphometry this taxon was previously reported as *M. triton* [50] and in most topologies it is also the sister taxon to the triton group. Genetically it is a very distinct lineage (13.5-14.4% K2P distance to taxa of the triton group) with a remarkably different karyotype than *M. triton* [52]. Earlier studies have already suggested that this taxon represents a valid species [51]. It can be sympatric with *M. mahomet*, but differs in habitat preferences; *M.* sp. "Harenna" lives mostly in the forests, while *M. mahomet* inhabits more open grassy habitats [[53]; L. Lavrenchenko, pers. obs.].

(2) The triton group (Figure 1b):

It is the group of MOTUs of relatively large body size, distributed mostly south of the equator (largely parapatric with the setulosus group - see Figure 1b vs. 1c). Genetic data suggest important cryptic variability (K2P distance among three MOTUs = 8.80-11.05%). Only nominotypical MOTU has a clear valid name, remaining lineages require further taxonomic studies.

(MOTU 4) *Mus triton* (Thomas, 1909)

This species was described from Mt. Elgon in Kenya and we provide the sequence from the type locality. It is distributed in the Kenyan highlands and northern part of Albertine rift. The same species probably occurs in southern Sudan also (described as *M. imatongensis*) [54]), but this should be confirmed by barcoding Sudanese specimens.

(MOTU 5) *Mus* sp. "Kikwit"

This distinct genetic lineage within the triton group was detected in two localities in south-western Democratic Republic of Congo (DRC). It may represent a new species, but more material and analyses are necessary to substantiate this claim. This MOTU supports important biogeographical distinctiveness of the Kikwit region in DRC (see also MOTU 21 from the minutoides group). The type locality of *Mus callewaerti* (Thomas, 1925) (Kananga, Kasaï occidental, DRC) is relatively near, so it is possible that they are conspecific, but a comparison with the type material will be necessary before a final conclusion can be reached (see also MOTU 6).

(MOTU 6) *Mus* cf. *callewaerti*

This taxon forms a well-supported separate lineage within the *triton* group. Its distribution range comprises a fairly important area situated between the Tanzanian Eastern Arc Mountains, through Southern Rift Mountains and northern Zambia till the Angolan highlands. In miombo woodlands of north-western Tanzania, it may have overlapping distribution ranges with *M. triton*, but no locality with sympatric occurrence was found in our study. The Angolan specimens were recently reported as *M. callewaerti* (Thomas, 1925) [14]. It is therefore possible that the whole clade should belong to *M. callewaerti*, but a comparison with type material will be necessary. The taxon prefers the miombo woodland or montane forest edges. There is important genetic variability within this taxon, with animals from Eastern Arc Mountains forming a distinct clade supported as a separate GMYC-species (Figure 3a).

(3) The setulosus group (Figure 1c):

We recognized five MOTUs within this highly supported monophyletic lineage. It includes relatively large-bodied species, with distribution ranges mostly north of the equator, i.e. largely parapatric with the triton group. Two of these MOTUs were only recorded in Ethiopia.

(MOTU 7) *Mus* cf. *proconodon*

It represents a lineage probably endemic to Ethiopia, where it mainly occurs in lowlands of the Rift Valley. We suggest assigning this MOTU to the species *M. proconodon* Rhoads, 1896, i.e. the Ethiopian taxon that was synonymised with *M. setulosus* [10] even if genetically it represents the most distinct lineage of the whole *setulosus* group.

(MOTU 8) *Mus setulosus* Peters, 1876

This highly supported MOTU from western-central Africa (north-west of the Congo River) represents the true *M. setulosus* (type locality is Victoria, Cameroon). The western border of its distribution likely lies in the dry region of the Dahomey gap. In the north-east (i.e. southern Central African Republic (CAR)), it is probably in contact with *M. bufo* (MOTU 10), and it is worthy of further study to analyse the possible contact zone and reproductive barriers between these two taxa in CAR.

(MOTU 9) *Mus* cf. *setulosus* "West"

MOTUs 8–11 form a monophyletic group of four strongly supported lineages with roughly parapatric distribution (Figure 1c). Two of them (MOTUs 8 and 9) have been previously named *M. setulosus* (e.g. [9]). MOTU 8 is distributed in central African forests, while MOTU 9 in western Africa (west of the Dahomey gap). MOTUs 10 and 11 represent valid species *M. bufo* (Thomas, 1906) and *M. mahomet* Rhoads, 1896, respectively. The topology and genetic distances (K2P distance = 8.1%) suggest that MOTUs 8 and 9 should be given different names. Because *M. setulosus* was described from Cameroon (i.e. distribution area of MOTU 8), we suggest that the West African populations of *M.* cf. *setulosus*, i.e. MOTU 9, may represent a separate new species, but this claim needs to be substantiated by further taxonomic work.

(MOTU 10) *Mus bufo* (Thomas, 1906)

The species was described from Ruwenzori Mts. in Uganda and it was considered endemic to the Albertine Rift. There are few sequences identified as *M. bufo* in GenBank. The first (Acc. no. DQ789905) from Bujumbura in

Burundi was reported by [9] as an incorrectly assigned species. Recently, new sequences of *M. bufo* from Kahuzi-Biega (DRC) were published [14] and all clearly cluster with the new sequences from CAR, DRC and Kenya reported in our study. Furthermore, we obtained a short sequence from the paratype of *M. bufo* from DRC (locality Idjwi) that also grouped with this clade. Although the morphological comparison with additional type material is necessary, we suggest that *M. bufo* has a much larger distribution range than previously assumed. This taxon may also involve additional populations of the setulosus group from Eastern Africa, especially those assigned to *M. emesi* Heller, 1911 (described from Uganda; morphologically similar to *M. mahomet*, with which it was synonymised [10]), and *M. pasha* Thomas, 1910 (East-African taxon that was synonymized first with *M. proconodon* and later on with *M. setulosus* [10]).

(MOTU 11) *Mus mahomet* Rhoads, 1896

It is an abundant species with a distribution range restricted to the Ethiopian Plateau. We provide the first sequences of this taxon, confirming its position within the setulosus group as a strongly supported monophyletic lineage. We therefore support the view of [55], who considered *M. mahomet* as an Ethiopian endemic, contrary to previous opinions merging it with Kenyan and Ugandan populations (i.e. most probably with MOTU 10, which is significantly supported sister group to *M. mahomet*; Figure 2).

(4) The baoulei group (Figure 1d):

This is a West African clade, until now known as a single species, but with very pronounced divergences between two subclades (mean K2P distance on *CYTB* = 9.46%) that have partially overlapping distribution ranges in Ghana and Ivory Coast. Only very limited genetic data are available, because the species of the baoulei group are probably rare or difficult to capture [12,13,23]. The species of this group occur in the forest-savannah ecotone and are generally larger than other West African species (except *M. setulosus*) [12]. The baoulei group is a sister lineage to the *sorella* group (Figure 2), which is also reflected in morphology [56].

(MOTU 12) *Mus baoulei* (Vermeiren & Verheyen, 1980)

The species *M. baoulei* was described from Lamto in the Ivory Coast [56]. Two individuals sequenced from the type locality [12] belong to the genetic clade that is distributed mainly in Ghana, Benin and western Nigeria (i.e. the type locality represents the westernmost record of this lineage).

(MOTU 13) *Mus* cf. *baoulei* "West"

Specimens from this lineage were found in Guinea and single individuals were sequenced from the eastern Ivory Coast [12] and Ghana [23]. Future more-detailed studies (using more samples, morphology and nuclear markers) are required to resolve whether MOTUs 12 and 13 represent separate species.

(5) The sorella group (Figure 1d):

It is a lineage of relatively large animals living in the Congo Basin's forest-savannah transit zones, but also reported from south-eastern Africa (Mozambique and Zimbabwe) [57]. While very limited genetic data are available, our sampling shows very divergent sequences that may represent up to four species, but more data are required for taxonomic revision of this group.

(MOTU 14) *Mus* sp. "Dakawa"

Two sequences from Dakawa (Tanzania) belong to the *M. sorella* group, but they are very distinct from other lineages of the group (K2P distance = 8.74-9.75%). It is possible that they represent a new species, but more taxonomic research is necessary. There is an existing name, *M. wamae*, that may be valid for this MOTU. This taxon was described as a member of the sorella group from the Kapiti Plains in southern Kenya [57].

(MOTU 15) *Mus* sp. "Koi River"

A single specimen from the moist savannah area near Koi River in south-western Ethiopia clearly belongs to the sorella group, but is very divergent at *CYTB* (K2P-distance between MOTU 15 and other lineages of the sorella group are 9.72-9.83%). Further taxonomic work is necessary to resolve the taxonomic rank of this lineage. This is the first record of the sorella group in Ethiopia.

(MOTU 16) *Mus sorella* (Thomas, 1909)

The first sequence of this MOTU was published under the name *M. sorella* by [58] from Sangba (CAR). The species *M. sorella* was described from hills around Mt. Elgon, an area which has clear biogeographical connections to CAR (see e.g. MOTU 10 or clade C of MOTU27; Figure 1c and f). We obtained one additional short sequence from this lineage by 454 pyrosequencing of a museum specimen from the Garamba National Park in north-eastern DRC, thus connecting Sangba with the type locality. However, it is also possible that these sequences represent another currently valid species described from CAR, i.e. *M. oubanguii* Petter & Genest, 1970 or *M. goundae* Petter & Genest, 1970. More samples and detailed analyses are required to resolve this taxonomic problem.

(MOTU 17) *Mus neavei* (Thomas, 1910)

Even if more morphological comparisons are necessary, hereafter we call this south-east African clade *M. neavei* and we report the first sequences of this species. The type locality of *M. neavei* (also morphologically belonging to the sorella group; [57]) is Petauke, Zambia. In our material, this taxon is distributed in hilly areas of southern Tanzania, Malawi and one locality in Zambia (not far from the type locality). It occurs in sympatry with MOTU 6

from the triton group [57] and in the Nyika Mountains in Malawi also with MOTU 1. The records from South African Republic (SAR) are not yet confirmed genetically; the specimen mentioned by [14] was finally identified as *M. minutoides* and no other sequences of *M. neavei* were obtained despite intensive recent sampling efforts in SAR (F. Veyrunes, pers. comm.)

(6)The minutoides group (Figures 1e-f):

This is the most diversified group within *Nannomys*, inhabiting various, mostly open habitats of sub-Saharan Africa. It harbours the real "pygmy" mice, i.e. the rodents with the smallest body size (some of them with body mass < 5 g). Most previous published genetic studies of *Nannomys* mainly targeted representatives of this group. Our phylogenetic analysis reveals three clear subgroups: subgroup 1 (MOTUs 18 to 20), subgroup 2 (MOTUs 21 and 22), and subgroup 3 (MOTUs 23 to 27).

(MOTU 18) *Mus* sp. "Zakouma"

A single specimen of this taxon was captured in the Zakouma National Park in south-eastern Chad [11]. It is genetically very distinct from its sister species, *M. mattheyi* F. Petter, 1969 and *M. haussa* (Thomas & Hinton, 1920), and further taxonomic work on more material from southern Chad may confirm it as a new distinct species. Together with *M. mattheyi* and *M. haussa*, this species forms a monophyletic group that diverged in West African savannahs.

(MOTU 19) *Mus haussa* (Thomas & Hinton, 1920)

It is a Sahelian taxon, recorded in the belt from Senegal to western Chad [9]. Similarly as in *M. mattheyi* and other West African savannah species of rodents [59-61], there is also indication of longitudinal genetic structure in *M. haussa*, but more detailed data are needed for more conclusive phylogeographical inferences.

(MOTU 20) *Mus mattheyi* F. Petter, 1969

M. mattheyi is typical species of Guinean savannah-forest mosaic from westernmost Africa (Senegal) to the Dahomey gap, the relatively dry region separating Guinean and Congolese forest blocks [9]. It is divided into western and eastern phylogeographic subclades with a presumable contact zone in the Ivory Coast (not shown). It is often the most abundant *Nannomys* in the rodent assemblages [13,23].

(MOTU 21) *Mus* cf. *kasaicus*

Two sequenced individuals from the Kikwit region (DRC) formed this genetically very distinct genetic MOTU. There are also indications from other rodent groups that the Kikwit area is a local centre of endemism (see e.g. MOTU 5 or [62]). There is an existing name, *M. kasaicus* (Cabrera, 1924), described from Kasaï Occidental Province, Kananga, DRC, for the taxon belonging morphologically to *M. minutoides* group [10], that may apply to this MOTU.

(MOTU 22) *Mus indutus* (Thomas, 1910)

M. indutus is a south African species, found in a relatively large area from northern Botswana to southern SAR [11,14,63,64]. Records from Zambia and Malawi are based on genotyping of old museum material [64] and should be taken with caution. It is probably sympatric with *M. minutoides* Smith, 1834 (= MOTU 27) in most of its distribution range.

(MOTU 23) *Mus* cf. *gratus*

Specimens from this taxon were typically captured in forest clearings and the ecotone between forest and open habitats in equatorial Africa. There are three distinct clades with clear west–east geographical structure: (i) a single specimen from lowland tropical forest in Congo (K2P distance to two remaining clades is cca 7%); (ii) the Kisangani region in DRC; and (iii) both montane and lowland tropical forests in southern Kenya and northern Tanzania. More taxonomic work is necessary to link this clade to an existing species; possibly *M. gratus* (Thomas & Wroughton, 1910), a taxon from the minutoides group described from eastern Ruwenzori, "upper Congo" and Virunga mountains. Again, the comparison with the types will be required to verify this hypothesis.

(MOTU 24) *Mus* cf. *gerbillus*

This taxon is distributed in dry Somali-Maasai savannah in Kenya and Tanzania. In all phylogenetic analyses, it is a sister clade to the Ethiopian MOTU 25 (mean K2P distance between these two clades is 8.87%). Further taxonomic work is necessary, but *M. gerbillus* (G.M. Allen & Loveridge, 1933) (currently the synonym for Tanzanian populations of *M. tenellus*) is an available name that may apply to this lineage.

(MOTU 25) *Mus* cf. *tenellus*

This lineage was found at two close localities in northern Ethiopia - in Hagere Selam and in the Mekelle University campus. It may represent true *M. tenellus* (Thomas, 1903) described from Blue Nile in Sudan, but the comparison with the type material is necessary. On the contrary, morphological studies of museum material suggested that most published Ethiopian records of *M. tenellus* were actually *M. minutoides* [10].

(MOTU 26) *Mus musculoides* Temminck, 1853

It is a typical species of the Sudanian savannah belt. It was previously reported from western Africa [11,12,17] and northern Cameroon [65]. We provide a new very distant record from western Ethiopia, representing the easternmost genetically confirmed locality of the species. Very probably it is also present in poorly sampled countries such as Chad, northern CAR and South Sudan.

(MOTU 27) *Mus minutoides* Smith, 1834

M. minutoides is a widely distributed species in most of sub-Saharan Africa (probably except continuously forested areas in the Congo Basin and deserts; Figure 1f). This MOTU also includes specimens from southern

Ethiopia; some of them were previously called *M. tenellus* [9,10]. This species has a very strong intraspecific phylogeographical structure. Median-joining network analysis of 131 sequences from this MOTU resulted in 84 haplotypes that form 11 strongly delimited haplogroups (Figure 4). The mean K2P-corrected distances among haplogroups ranged from 1.21% (TZw vs. KE) to 3.65% (ZA vs. Chin). All haplogroups are connected in the form of a star, suggesting multiple synchronous vicariance events. Allopatric divergences with subsequent expansions are further supported by current parapatric distribution of most clades and frequent, but narrow, secondary contacts among them (Figure 1f). The geographic structure within individual haplogroups is relatively weak, except clade SE, where it is possible to distinguish the separate sublineages from South Africa (h79-h81), Mozambique (h15-h17) and Tanzania (remaining haplotypes). Two haplogroups are only represented by animals from single localities (Minziro and Chingombe in Tanzania), but it is possible that they are more widespread in neighbouring regions in eastern DRC, where the relevant samples are missing.

Divergence dating

The basal split of the extant *Nannomys* was dated at 5.24 Mya with 95% of the highest posterior density (HPD) between 4.58–5.96 Mya. Successive divergence of the extant major species groups then took place throughout Pliocene, with median estimates of divergence times ranging from 4.9 Mya (split off of MOTU 1 "Nyika") to 2.44 Mya, i.e. the divergence of MOTU 23 (cf. *gratus*) and MOTUs 24–27 (i.e. four other species of the minutoides group) (Additional file 2). Posterior estimates of divergence dates at the calibration points are shifted towards past in the case of *Apodemus* (prior median 5.89, posterior median 7.38) and *Arvicanthis-Otomys* (6.81 vs. 8.13) but towards the present in the case of *Mus* (8.00 vs. 7.44). Two other divergence dates are also worth noting: the split-off of *Myomyscus yemeni* estimated at 6.21 (5.12–7.33) Mya, which is consistent with its migration to Arabian peninsula across the land bridge during the Messinian crisis, and the origin of modern *Otomys* 3.77 (2.83–4.81) Mya, first appearing in the fossil record around 3 Mya [[66], p.290]. Complete results of the

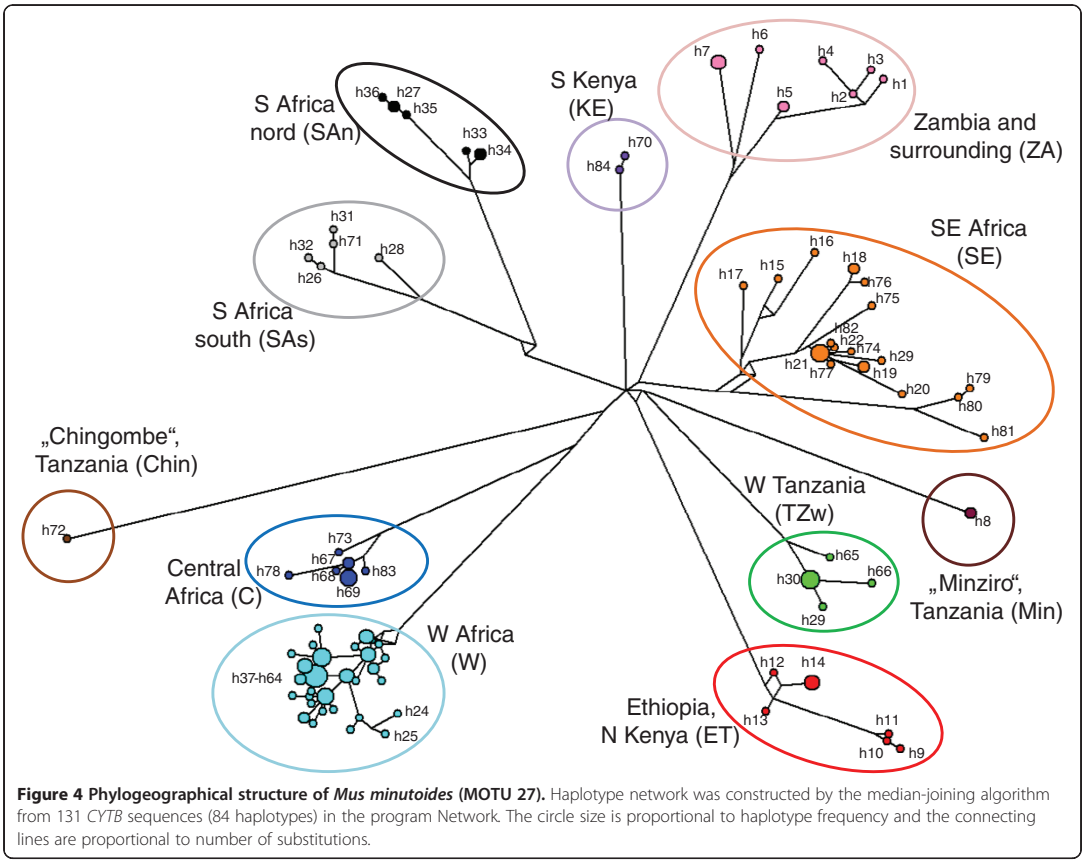


Figure 4 Phylogeographical structure of *Mus minutoides* (MOTU 27). Haplotype network was constructed by the median-joining algorithm from 131 *CYTB* sequences (84 haplotypes) in the program Network. The circle size is proportional to haplotype frequency and the connecting lines are proportional to number of substitutions.

divergence dating analysis are reported in Additional file 2.

The full set of branching times between 27 MOTUs is given in Figure 3a. It is based on the secondary dating of the ultrametric tree for GMYC, but the posterior estimates of divergence dates are consistent with previous analysis (compare Figure 3a and Additional file 2). Main species lineages diverged in lower Pliocene (5.2-4.5 Mya) and an intensive period of speciation is also visible in the lower Pleistocene (2.1-1.6 Mya), when many extant lineages within main species groups appeared.

Biogeographical analysis

Bayesian analysis of discrete traits in BEAST revealed that the most ancestral distribution (98% support) of *Nannomys* included mountains of Eastern Africa (Figure 3b). This type of distribution is currently present in all three ancient monotypic lineages (MOTUs 1–3), as well as in numerous lineages of the triton and setulosus groups. There are two major habitat shifts in the *Nannomys* evolution. (1) The lineage leading to the baoulei group colonized the forests (and forest-savannah mosaic) in western Africa cca 4 Mya, where it split to western and eastern sublineages later on; (2) the minutoides group descended from mountains, adapted to more arid open habitats, and started to radiate across the whole sub-Saharan Africa cca 3.5 Mya. In the first radiation phase, MOTUs 18–27 speciated in savannah-like habitats over all of Africa (approx. 3.5-1.6 Mya). Geographically similar, but much more recent (cca 1 Mya) radiation occurred inside MOTU 27, i.e. *M. minutoides* (Figures 1f, 3a, and 4).

Very similar results were obtained by the maximum likelihood approach in Lagrange (Additional file 5). Most basal splits occurred with the highest probability in the mountains of East Africa, also where most of the MOTUs from the triton group diverged. The first clear shifts to other habitats are visible in the ancestors of the baoulei group (to the forests or forest edges, where both MOTUs from this group occur until today) and in the ancestors of the minutoides group (to the savannah). The most intensive radiation in the latter took place in savannahs, with one shift to the forest habitat detected in MOTU 23 (*M.* cf. *gratus*). The estimates of ancestral ranges are less clear in the setulosus and the sorella groups. While the former started to diverge most probably in mountains (with subsequent spreading of two "*setulosus*" MOTUs to forests of central and eastern Africa), the latter had ancestors occurring with similar support either in savannahs or in hills of Eastern Africa.

Discussion

For the purpose of our study, we compiled new and existing sequences into the largest genetic dataset to date of the subgenus *Nannomys* and performed the first phylogenetic analysis of the group that contains most of the currently recognized valid species across the whole sub-Saharan Africa. We detected a surprisingly high amount of cryptic diversity, with numerous candidates for new species. Wide geographical sampling also allowed the first empirical definition of the distribution areas of all the detected lineages based on physically present genotyped individuals. Using several calibration points and the current distributional data, we also carried out biogeographical analysis and reconstructed the possible evolutionary scenario of this highly successful group of sub-Saharan murines.

Species concepts and estimation of the number of *Nannomys* species

Species diversity crucially depends on the adopted species concept. Widely used concepts of typological or biological species are not always applicable for species delimitation because of frequent convergent evolution, cryptic species, and the impossibility of proving reproductive isolation among allopatric populations. Together with the rapidly increasing amount of genetic data from free-living populations, these concepts are often complemented by genetic [37] or phylogenetic [3] species concepts, creating the so-called integrative taxonomic approach. Although genetic approaches can sometimes lead to an unjustified increase in the number of species (so-called taxonomic inflation [2,67]), they often detect cryptic diversity within evolutionary lineages that can be generally important from the taxonomic as well as conservation point of view. In our study, we used the combination of maximum likelihood delimitation of phylogenetic species and the genetic distances to estimate the number of MOTUs (= putative species) of *Nannomys* in Africa. We are aware of the drawbacks of these approaches (e.g. the use of only maternally inherited mtDNA), however, our aim was not to perform the taxonomic changes based solely on limited genetic data, but rather to identify the taxa and regions of high cryptic diversity requiring more detailed taxonomic studies.

The combination of different approaches revealed the existence of 27 MOTUs. This is considerably more than the 18 currently accepted *Nannomys* species [9,10], suggesting that numerous putative species have so far remained undetected, and therefore undescribed. Most of the genetic data of *Nannomys* that have been collected to date originate from Western and Southern Africa, where the taxonomy of this group has been intensively explored (reviewed by [9]). The number of candidates for potential new species in western Africa revealed by our study is therefore relatively low (only *M.* cf. *setulosus* "West" or *M.* cf. *baoulei* "West") and it is also possible that these MOTUs just represent marked phylogeographical structure of within-species lineages with parapatric distribution

(but see [12] that already suggested *M. setulosus* as a species complex).

The situation is completely different in Eastern Africa, from where only fragmentary genetic data were available prior to this study. Our results may lead to the description of more than 10 new species that are already now sufficiently delimited using the combination of genetic, ecological and geographic data. Many of these so far undescribed taxa occur in mountains or highland habitats, but a few other potential new species (like *M.* cf. *gerbillus*) are typical inhabitants of dry savannahs. The taxonomic diversity of *Nannomys* is probably the highest in Ethiopia. As for many other organisms, the Ethiopian highlands represent an important hot-spot of African endemism for *Mus*. We have revealed the presence of 8 MOTUs in this country, and only two of them (*M. minutoides* and *M. musculoides*) have also been recorded outside Ethiopia. The six remaining species are probably endemic, making *Mus* the genus with the second highest number of Ethiopian mammal endemics (after *Lophuromys* with 9 endemic species; [68]).

Even if we have not sequenced the type material of most currently valid taxa (except paratypes of *M. bufo*), we have been able to assign the most probable species names to 13–14 MOTUs based on previous genetic studies (including karyotypes; [11,12]), geographical distribution (i.e. sequences from the type locality or close neighbourhood) and external morphology. Therefore, the genetic dataset from this study represents a solid basis for future identification of morphologically similar *Nannomys* species via DNA barcoding (using e.g. evolutionary placement algorithm; [69]). Unfortunately, our dataset lacks sequences of four valid species. *M. oubanguii* Peter & Genest, 1970 and *M. goundae* Peter & Genest, 1970 represent two species from the sorella group known only from few localities in the Central African Republic. They were described mainly on the basis of external morphology [57] and their specific status has been questioned previously ([10]; but see conspicuous differences in karyotypes of these two species - reviewed in [9]). The whole sorella group requires a profound revision including new sampling in savannahs north of the Congo Basin and additional genetic data. We found high genetic variation within the sorella group, but most clades are represented by only one or two localities (except *M. neavei*) and in most cases it is not possible to assign the particular clades to currently valid species names. *M. setzeri* Petter, 1978 is a rare taxon with limited distribution in dry areas of Namibia, Botswana and western Zambia [70,71]; it is probably a valid species as it can be morphologically distinguished from sympatric *Nannomys* species [63,72]. *M. orangiae* Roberts, 1926 is the fourth species that is currently valid and missing from our dataset. It also is a southern African species with unclear taxonomical status. It was previously considered a subspecies of either *M. setzeri* or *M. minutoides* [10] and may just represent one of the cytotypes of the latter [9].

Phylogenetic estimate of species richness of *Nannomys* in our study (25–30 MOTUs that may represent separate species) suggests that it is one of the most speciose groups of African terrestrial mammals. Similarly well studied species-rich genera of African rodents usually have a lower number of monophyletic genetic lineages considered as species, e.g. *Praomys* (16–20 species; [73]) or *Hylomyscus* (21 species, including undescribed and recently described taxa; [74], J. Kennis et al., submitted). The only genus with higher described species richness is *Lophuromys* (29 species; [10,68,75]). However, this genus is specialized to tropical forests and ecotones and it is likely that intensive genetic drift in fragmented forest habitats (especially in Eastern Africa) caused morphological distinctiveness allowing differentiation of a high number of genetically similar morphospecies [68,75]. It is also worth to note that in comparison with the above-mentioned genera, *Nannomys* colonized a much wider spectrum of habitats (from Afroalpine meadows and mountain forests to very arid savannah).

Mus minutoides as a model for pan-African phylogeography

The MOTU with the largest distribution of all *Nannomys* is *M. minutoides* (=MOTU 27). There are only a very few such widespread savannah-forest mosaic species distributed across almost complete sub-Saharan Africa. Among rodents, only the ubiquitous *Mastomys natalensis* had held this habitat breadth, and it was considered the rodent species with the largest distribution area in Africa [10]. Our genetic data confirm that *M. minutoides* has very similar and probably even larger distribution than *M. natalensis*. It can be argued that MOTU 27 does not represent a single species but rather a species complex, which may be supported by the GMYC analysis revealing significant support for additional speciation events within this clade (see Figure 3a). However, in absence of more detailed evidence, we prefer to maintain all genetic lineages of MOTU 27 within the species *M. minutoides*. They do not show visible external differences (although detailed morphological analysis of genotyped material is still missing), they radiated relatively recently (last 1 Mya) and the Tamura-Nei corrected genetic distances among clades (1.21-3.65% on *CYTB*) are comparable with those among clades of *M. natalensis* (2.1-3.8%; [76]), i.e. much lower than usual genetic distances between sister species of rodents [37]. Further detailed studies should focus on the contact zones of divergent clades to reveal whether they can interbreed or not.

Species with large distributions and strong affinities to open habitats can serve as possible models for comparative pan-African phylogeography of the savannah-like biomes.

Recent phylogeographic studies of *M. natalensis* showed that populations were strongly influenced by Pleistocene climate fluctuations [76]. The presence of genetically divergent clades with parapatric distribution is congruent with the scenario invoking allopatric fragmentation and vicariance. Almost the exact same geographic pattern of genetic differentiation is visible in *M. minutoides* (compare Figure 1f in this study with Figure 1 in [76]). The phylogeographic pattern suggests at least 11 different savannah refugia approximately 1 Mya, i.e. in the period of very strong climatic instability [77]). The genetic lineages evolved in allopatry and subsequently spread during suitable periods of savannah expansion. Further research should focus on precise localization of refugia by combining information from population genetics with modelling of past ecological conditions [78]).

A new biogeographical scenario of *Nannomys* radiation in Africa - from mountains to lowland forests, savannahs and arid Sahelian environments

More complete taxon sampling from the whole sub-Saharan Africa now allows significant modification and increased precision of the previously proposed biogeographical scenario of *Nannomys* radiation in Africa [11]. Our molecular dating based on plausible paleontological calibration and taxon-unbiased phylogenetic tree suggests that the divergence of the genus *Mus* to the current subgenera occurred in Asia in the late Miocene (cca 6.8-7.4 Mya), which is in good agreement with previous studies [7,11,44]. The colonization of Africa by *Mus* occurred very probably in the Messinian period (7.3-5.3 Mya) when the temporary land bridge connected Africa and southwest Arabia. In this period, many faunal exchanges between Africa and Asia are well documented [79-81]. It is therefore highly probable that *Mus* was already in Africa at the beginning of the Pliocene. The basic split of the extant *Nannomys* was dated at 5.24 Mya (95% HPD 4.58–5.96 Mya), i.e. very soon after a land bridge between Africa and Southwest Arabia disappeared (5.3-6 Mya; [82,83]). The oldest fossil evidence of the genus *Mus* in Africa was from the early to middle Pliocene in Ethiopia (the Omo valley in the south of the Ethiopian Rift Valley and Hadar in the east, 5–2.5 Mya; [84,85]) and Kenya (4.5 Mya; [27]).

Due to incomplete sampling (mainly in eastern Africa) previous studies could not adequately explore the evolutionary history of *Nannomys*, especially since our biogeographical reconstructions demonstrate that the first divergence of *Nannomys* occurred in eastern Africa. Paleoclimatic and paleoanthropological research in eastern Africa suggested repeated association of critical events in hominin evolution with the most prolonged intervals of high climate variability. Potts (2013) [77] defined eight intervals of predicted high climate variability in the last 5 My and argued that most important events in hominin evolution occurred within these periods. Three of the most prolonged intervals of predicted high climate variability are 2.79-2.47 Ma, 1.89-1.69 Ma, and 1.12-0.92 Ma and they largely overlap with the previously defined periods of the occurrence of large lakes [86] as well as with inferred aridity phases based on dust records, paleosol δ13C, and the prevalence of grazing bovids [87].

Clear associations between periods of climatic instability and divergence events are also visible in phylogenetic reconstructions of *Nannomys*. The first splits leading to ancestors of most current species groups probably occurred in eastern Africa in the period 5.2-4.5 Mya (Figure 3a), which corresponds to the longest era of strong wet-dry variability [77]. Nothing is known about the ecology of the extinct *Mus* taxa, but surviving ancient lineages (MOTUs 1–3) may provide some clues. They can be considered "living fossils", i.e. monotypic relict taxa living in very restricted areas in Eastern African mountains. The period 4–3.5 Mya is considered relatively stable with few documented evolutionary events [77] and we observed only two vicariance events in *Nannomys* during this period. The first is the north–south split of MOTU 3 (*M.* sp. "Harenna") and the triton group, and the second is the west–east split of the baoulei and the sorella groups (see Figure 3a and compare it with distributions at Figure 1). The most intensive radiation of *Nannomys* is dated into the period 3.5-1.4 Mya (see Figure 3a), when most current MOTUs (i.e. putative species) appeared. The beginning of this period coincides with the start of a cooling and aridification trend [88]. The open savannah-like habitats were spreading intensively and at the same time the climate was very variable (four prolonged periods of strong wet-dry variability are dated into this range; [77]). This variable climate likely yielded environmental changes that increased the frequency of evolutionary responses like adaptation, dispersal (especially in open habitats), and ultimately, speciation (for example it was also the period with the highest number of hominin taxa; [89]). Our biogeographic analyses are consistent with these findings because the most intensive radiation occurred in the minutoides lineage in savannahs. The presumed shift from mountains to more arid and open habitats was clearly linked with the decrease of the body size in the minutoides lineage. The ancient *M. imberbis* (MOTU 2) has a body size of 25 g [8], MOTU 1 (*M.* sp. "Nyika") has 14 g, MOTU 3 (*M.* sp. "Harenna") has cca 16 g (our unpublished data) and the members of other non-*minutoides* groups weight 8–13 g [9]. In contrast, all species of the *minutoides* clade have body size 3–8 g, making them one of the smallest mammals in the world [9]. The last period of climatic instability is dated to 1.12-0.92 Mya, which coincides with the likely simultaneous split of the MOTU 27, i.e. *Mus minutoides*, into 11 distinct genetic lineages (see above).

Ecological constraints and multi-species sympatry

Previous studies revealed that at several sites more than one species of *Nannomys* occurs in sympatry [12,13,23]. Their observations are in agreement with the distribution ranges based on genotyped individuals (Figure 1) showing largely overlapping distribution areas of many species. However, if we exclude widely distributed *M. minutoides* (MOTU 27), the distribution of individual species within the same species group is predominantly parapatric (most illustrative in Figure 1b, c, d), while sympatry is typical for species from different species groups. This suggests that the species groups might have evolved specific morphological adaptations that allow their sympatric occurrence with the members of other *Nannomys* lineages. Although detailed morphological analysis of genetically identified specimens is still missing, preliminary data suggest clear differences in the skull morphology among the species groups ([12], E. Verheyen et al., unpublished data), with possible functional consequences in separation of ecological niches (for example dietary).

Even if the distribution areas of two or more species from the same species group overlap, closer examination of our data provide evidence for the preference of different habitats. For example two Ethiopian endemics from the setulosus group, *M.* cf. *proconodon* (MOTU 7) and *M. mahomet* (MOTU 11), have never been captured at the same locality; the former prefers lowland habitats in the Rift Valley, while the latter is common species across the Ethiopian highlands. Up to four species of the minutoides group can be found in western Africa, but their ecological requirements are probably different. Based on the data summarized at Figure 1 and published records, it seems that *M. minutoides* is able to live in Western Africa in relatively humid places, *M. mattheyi* prefers dry Guinean savannah and the transition zone between forest and savannah, *M. musculoides* is a typical inhabitant of Sudanian savannah belt from Guinea to western Ethiopia and *M. haussa* lives in arid Sahelian environment ([12,65], figure one in this study). Similarly *M. minutoides* can occasionally be found in the same localities as *M. indutus* in southern Africa, but the latter probably prefers drier habitats ([63,64] and references therein).

Relevance to the understanding of karyotype evolution and sex determination

The subgenus *Nannomys* has previously been used as a suitable model for studies of karyotype evolution due to very high variability of chromosomal rearrangements [11,17,90,91]. The ancestral karyotype of the pygmy mice was composed of 36 acrocentric chromosomes [17,92], but the wide spectrum of mutational mechanisms modified the chromosomal constitution. Besides relatively frequent Robertsonian translocations, other chromosomal rearrangements were described in *Nannomys*, including variable sex-autosome translocations, pericentric inversions, tandem fusions and WARTs (Whole-Arm-Reciprocal Translocation) [9]. Pan-African phylogeny based on more complete taxon sampling presented in our study can help to understand the karyotype evolution in general and sex determination mechanisms in particular. The mapping of karyotype features on the phylogenetic tree can help to define specific predictions that can be further verified by sampling focussed on particular species and geographical areas.

For example, tandem fusions - one of the rarest chromosomal rearrangements - were evidenced in *M. triton* (MOTU 4) and *M.* sp. "Harenna" (MOTU 3) that in most phylogenies cluster together. Even if they were suspected in two other species of the sorella group in the CAR (*M. goundae* and *M. oubangui*, not sampled in our study; [9]), further detailed studies of these rare mutations should direct their focus on widely distributed and common MOTU 6 belonging to the triton group. One of the most conspicuous features of *Nannomys* karyotypes is the fusions of autosomes and sex-chromosomes. These fusions were most frequently studied in two terminal taxa of the minutoides group (MOTU 26 - *musculoides* and MOTU 27 - *minutoides*), but they were also observed in *M. goundae* and *M. oubangui* (very probably belonging to the sorella group) and in *M. triton* (MOTU 4) [9]. Since they appeared several times independently, it is therefore clear that predispositions for translocations of sex chromosomes exist in more lineages of *Nannomys*, yet these translocations are not a general feature of the whole subgenus, as, for example the setulosus group is very conservative and all MOTUs karyotyped until today have the ancestral karyotype (2n = 36, NF =36) ([78] and references in [9]). Future research on East African species of the minutoides group (MOTUs 23–25, i.e. *M.* cf. *tenellus*, *M.* cf. *gerbillus*, and *M.* cf. *gratus*), the sorella group and the triton group could thus potentially bring interesting new insights on the evolution and polymorphism of sex-autosome translocations. Finally, the phylogeographic pattern described in our study for the most karyotypically variable species, *M. minutoides*, can help to design further sampling of chromosomal data in lineages, where the karyotypes are not yet known. The haplotype network suggests 11 main lineages that probably differentiated in small allopatric populations at the same time, which could have led to establishment and fixation of important karyotypic differences [64,90]), possibly involving presently unknown means of sex determination [91]. If such karyotype differences among genetic lineages exist, it would be also extremely interesting to study the possible contact zones among them (see Figure 1f).

Consequences for future epidemiological studies

Rodents are reservoir hosts of important human pathogens, of which some can cause serious diseases. Most

recent examples of emerging and re-emerging diseases have been caused by RNA viruses [93] and understanding of their evolution and epidemiology is essential for predicting future emergences and designing interventions (e. g. vaccinations). Among RNA viruses hosted primarily by African rodents, the Lassa arenavirus has received most attention, because it is responsible for Lassa hemorrhagic fever in West Africa, which causes thousands of human deaths each year [94]. The host specificity of arenaviruses is thought to be relatively strict with often a single species described as the primary reservoir host [21]. A long-term evolutionary history between arenaviruses and their hosts (co-evolution) was originally suggested due to the almost perfect sorting of arenavirus lineages into rodent clades (e.g. [95]). Recent studies suggest that pygmy mice are frequent hosts of arenaviruses (and probably also other important parasites) and they often live close to human habitations (e.g. [13]). The pan-African phylogeny of *Nannomys* proposed in this paper can help to describe the co-evolutionary patterns of arenaviruses and their hosts and even provide the potential clues for understanding the occasional disease outbreaks.

The first virus found in *Nannomys* was the virus Kodoko, described from *Mus minutoides* in western Africa, and belonging to the lineage of lymphocytic choriomeningitis virus (LCMV) that is hosted by the house mouse [20]. This finding (followed by description of new Kodoko strain in Eastern African, [21]) thus supported a co-evolutionary scenario, because all arenaviruses known from African murine hosts at that time grouped according to taxonomic position of theirs hosts (i.e. three main lineages of African arenaviruses were hosted by rodents of three tribes, Praomyini, Arvicanthini and Murini; Lecompte et al. 2007, Gouy de Belocq et al. 2010). However, the next arenavirus, called Gbagroube and described from *Mus* cf. *setulosus* (MOTU 9 in this study) from the Ivory Coast, does not belong to the LCMV lineage (specific to *Mus*), but surprisingly clusters with the Lassa virus strains [22]. Very recently, two other arenaviruses were found in *Nannomys* in Ghana [23]. The virus Natorduori is hosted by *M. mattheyi* (MOTU 20, the minutoides group) and clusters clearly into *Mus*-specific LCMV lineage. In contrast, the virus Jirandogo, the first arenavirus reported from *M. baoulei*, in various phylogenies based on its different genome segments belongs to the Lassa virus group (similarly as Gbagroube virus). African pygmy mice are therefore the first group of African rodents that host two very different lineages of arenaviruses; one of them seems to be *Mus*-specific (in Africa now reported from two species in the minutoides group), but the second forms the sister lineage of the highly pathogenic Lassa virus (hosted by species from the setulosus and the baoulei groups). Further surveillance for new arenaviruses focussed preferentially on *Nannomys* lineages where no viruses have yet been found (e.g. the triton or sorella groups widely distributed in central and eastern Africa) can increase understanding of the evolution of these pathogens and predict the regions of possible epidemiological importance.

Conclusions

The known species diversity of tropical organisms is highly underestimated even for relatively well known animals like mammals. Here we performed a phylogenetic analysis of the largest available set of genetic data collected from the only indigenous African lineage of the genus *Mus*, called *Nannomys*. A conservative definition of MOTUs suggests that the number of species described to date represents only approximately 60% of possible species diversity and intensive taxonomic work is now required to allow the formal description of genetically divergent lineages. We also provide the first reliable genotype-based distribution ranges of particular MOTUs that can aid in future species inventories in different parts of Africa. The dating of divergences and biogeographical analyses strongly suggest that ancestors of *Nannomys* colonized Africa at the end of Miocene and diverged to ancestors of the main species groups in mountains of Eastern Africa in lower Pliocene. The aridification that started in Africa cca 3 Mya led to spreading of open habitats and provided new ecological niches that were fully utilized by *Nannomys*. In particular, the so-called minutoides lineage underwent an exceptionally intensive radiation in savannah-like habitats and occupied almost whole sub-Saharan Africa in several colonization waves. The combination of a detailed phylogeny based on an almost complete taxon sampling combined with genotype-based distributional data of lineages, taxa and valid species provides a solid foundation to address specific ecologically-explicit evolutionary hypotheses using *Nannomys* as a model system, i.e. in evolution of sex determination and host-virus co-evolution.

Availability of supporting data

The newly produced sequences were submitted to GenBank under accession numbers KJ935741-KJ935873 (*CYTB*) and KJ935874-KJ935905 (*IRBP*) (see Additional file 1 for more details). The final alignment of concatenated sequences used in phylogenetic analyses is in Additional file 6.

Additional files

Additional file 1: Details on collecting localities and genetic data for all *Nannomys* specimens.

Additional file 2: List of sequences used for divergence dating and resulting fossil-calibrated timetree.

Additional file 3: Maximum likelihood phylogeny of *Nannomys* based on separate analyses of mitochondrial *CYTB* and nuclear *IRBP* genes.

Additional file 4: Distribution of genetic distances at *CYTB* within and among taxa delimited by different methods.

Additional file 5: Biogeographical reconstruction of ancestral distribution of *Nannomys* lineages using maximum likelihood in Lagrange.

Additional file 6: Alignment of 179 ingroup and 7 outgroup concatenated sequences of *CYTB* and *IRBP*.

Abbreviations

CYTB: Mitochondrial gene for cytochrome *b*; IRBP: Gene for interphotoreceptor binding protein; Mya: Million years ago; mtDNA: Mitochondrial DNA; ML: Maximum likelihood; BI: Bayesian inference; MOTU: Molecular operational taxonomic unit; GMYC: Generalized mixed Yule-coalescent model; PCR: Polymerase chain reaction; AIC: Akaike information criterion; LCMV: Lymphocytic choriomeningitis virus.

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

JB, RŠ and EV conceived and designed the study, JB, RŠ, VM, YM, LL, KW, NO, JM, NA, MC, HL collected important part of samples, TA and YM performed laboratory analyses, JB and OM analysed data (phylogenetic and biogeographic analyses), and JB (with help of OM) wrote the first draft of the manuscript. All authors contributed to the final version of the paper. All authors read and approved the final manuscript.

Authors' information

JB is head of molecular ecology group at Institute of Vertebrate Biology ASCR, generally interested in factors affecting evolution of vertebrate populations. His actual topics include phylogeography and speciation in Africa, conservation genetics and mechanisms of host-parasite co-evolution. OM and TA are post-doc and PhD student in JB's lab, respectively. RŠ, YM, LL, HL and EV are leaders of research groups studying ecology and evolution of vertebrates, mainly mammals and especially rodents, in different parts of Africa. VM, NO, JM, KW, NA and MC are African and European collaborators of above-mentioned researchers, with joint interest in biodiversity of African rodents.

Acknowledgements

This study was supported by the projects of the Czech Science Foundation, no. P506-10-0983 (JB and RŠ), the Russian Foundation for Basic Research, no. 12-04-01283 (LL), and the Belgian Science Policy project 'Evaluating the effect of Pleistocene climate changes on speciation patterns in selected African vertebrates' (EV). For help in the field, we acknowledge H. Patzenhauerová-Konvičková, J. Šklíba, M. Lövy, C. Sabuni, G. Mhamphi, F. Sedláček, S. Šafarčíková, A. Konečný, S. Gambalemoke Mbalitini, and all local collaborators. The assistance of R. Makundi, W.N. Chitaukali, B. Dudu Akaibe and late W. Verheyen with project logistics and sample collection is highly appreciated. H. Patzenhauerová-Konvičková and L. Piálek helped with genotyping and E. Fichet-Calvet and S. Gryseels provided unpublished sequences. For permission to carry out the research and to collect specimens we are obliged to the National Research Council and Forestry Department in Malawi, the National Council for Science and Technology, the Kenyan Forest Service and the Kenyan Wildlife Service, the Ethiopian Wildlife Conservation Authority, Zambian Wildlife Authority, Sokoine University of Agriculture, and the 'Centre de surveillance de la biodiversité' in Kisangani, RD Congo. We would also like to thank the curators of the museums (G. Csorba, N. Duncan, V. Nicolas and W. Wendelen) for allowing us to study the material in their care, and two anonymous reviewers and J.C. Winternitz for comments on the previous version of the manuscript.

Author details

[1]Institute of Vertebrate Biology, Academy of Sciences of the Czech Republic, Brno, Czech Republic. [2]Department of Botany and Zoology, Faculty of Science, Masaryk University, Brno, Czech Republic. [3]Institute of Animal Physiology and Genetics, Academy of Sciences of the Czech Republic, Brno, Czech Republic. [4]Department of Zoology, Faculty of Science, University of South Bohemia, České Budějovice, Czech Republic. [5]Department of Biology, College of Natural and Computational Sciences, Mekelle University, Tigray, Ethiopia. [6]A.N.Severtsov Institute of Ecology and Evolution RAS, Moscow, Russia. [7]Earth Watch Institute, Nairobi, Kenya. [8]College of Agriculture and Veterinary Sciences, University of Nairobi, Nairobi, Kenya. [9]University of Kisangani, Eastern Province, Kisangani, DR, Congo. [10]CNRS UMR 6552/53, Université de Rennes 1, Station Biologique, Paimpont, France. [11]Evolutionary Ecology Group, Biology Department, University of Antwerp, Antwerpen, Belgium. [12]Royal Belgian Institute for Natural Sciences, Operational Direction Taxonomy and Phylogeny, Brussels, Belgium. [13]Institute of Vertebrate Biology, Academy of Sciences of the Czech Republic, Research Facility Studenec, Studenec 122, 675 02 Koněšín, Czech Republic.

Received: 30 May 2014 Accepted: 27 November 2014
Published online: 14 December 2014

References

1. Mora C, Tittensor DP, Adl S, Simpson AGB, Worm B: **How many species are there on earth and in the ocean?** *PLoS Biol* 2011, **9**(8):e1001127.
2. Zachos FE, Apollonio M, Bärmann EV, Festa-Bianchet M, Göhlich U, Habel JC, Haring E, Kruckenhauser L, Lovari S, McDevitt AD, Pertoldi C, Rössner GE, Sánchez-Villagra MR, Scandura M, Suchentrunk F: **Species inflation and taxonomic artefacts - A critical comment on recent trends in mammalian classification.** *Mamm Biol* 2013, **78:**1–6.
3. Groves CP: **The nature of species: A rejoinder to Zachos et al.** *Mamm Biol* 2013, **78:**7–9.
4. Krishna Krishnamurthy P, Francis RA: **A critical review on the utility of DNA barcoding in biodiversity conservation.** *Biodivers Conserv* 2012, **21:**1907–1919.
5. Hoffmann M, Grubb P, Groves CP, Hutterer R, Van der Straeten E, Simmons N, Bergmans W: **A synthesis of African and western Indian Ocean Island mammal taxa (Class: Mammalia) described between 1988 and 2008: an update to Allen (1939) and Ansell (1989).** *Zootaxa* 2009, **2205:**1–36.
6. Macholán M, Baird SJE, Munclinger P, Piálek J: *Evolution of the house mouse.* New York: Cambridge University Press; 2012.
7. Chevret P, Veyrunes F, Britton-Davidian J: **Molecular phylogeny of the genus *Mus* (Rodentia: Murinae) based on mitochondrial and nuclear data.** *Biol J Linnean Soc* 2005, **84:**417–427.
8. Meheretu Y, Šumbera R, Bryja J: **Enigmatic Ethiopian endemic rodent *Muriculus imberbis* (Rüppell 1842) represents a separate lineage within genus *Mus*.** *Mammalia* 2014, in press (doi:10.1515/mammalia-2013-0119).
9. Britton-Davidian J, Robinson TJ, Veyrunes F: **Systematics and evolution of the African pygmy mice, subgenus *Nannomys*: A review.** *Acta Oecol* 2012, **2012**(42):41–49.
10. Musser GG, Carleton MD: **Superfamily Muroidea**. In *Mammal Species of the World: A Taxonomic and Geographic Reference*. Edited by Wilson DE, Reeder DM. Baltimore: Johns Hopkins University Press; 2005:894–1531.
11. Veyrunes F, Britton-Davidian J, Robinson TJ, Calvet E, Denys C, Chevret P: **Molecular phylogeny of the African pygmy mice, subgenus *Nannomys* (Rodentia, Murinae, *Mus*): implications for chromosomal evolution.** *Mol Phylogenet Evol* 2005, **36:**358–369.
12. Kan Kouassi S, Nicolas V, Aniskine V, Lalis A, Cruaud C, Couloux A, Colyn M, Dosso M, Koivogui L, Verheyen E, Akoua-Koffi C, Denys C: **Taxonomy and biogeography of the African Pygmy mice, Subgenus *Nannomys* (Rodentia, Murinae, *Mus*) in Ivory Coast and Guinea (West Africa).** *Mammalia* 2008, **72:**237–252.
13. Fichet-Calvet E, Audenaert L, Barrière P, Verheyen E: **Diversity, dynamics and reproduction in a community of small mammals in Upper Guinea, with emphasis on pygmy mice ecology.** *Afr J Ecol* 2009, **48:**600–614.
14. Lamb J, Downs S, Eiseb S, Taylor PJ: **Increased geographic sampling reveals considerable new genetic diversity in morphologically conservative African Pygmy Mice (Genus *Mus*; Subgenus *Nannomys*).** *Mamm Biol* 2014, **79:**24–35.
15. Matthey R: **L'eventail robertsonien chez les *Mus* (*Leggada*) africains du groupe *minutoides-musculoides*.** *Rev Suisse Zool* 1970, **77**(3):625–629.
16. Jotterand M: **Un nouveau système polymorphe chez une nouvelle espèce de *Leggada* (*Mus goundae* Petter).** *Experientia* 1970, **26:**1360–1361.
17. Veyrunes F, Catalan J, Sicard B, Robinson TJ, Duplantier JM, Granjon L, Dobigny G, Britton-Davidian J: **Autosome and sex chromosome diversity among the African pygmy mice, subgenus *Nannomys*.** *Chromosome Res* 2004, **12:**369–382.
18. Veyrunes F, Watson J, Robinson TJ, Britton-Davidian J: **Accumulation of rare sex chromosome rearrangements in the African pygmy mouse, *Mus* (*Nannomys*) *minutoides*: a whole-arm reciprocal translocation (WART) involving a X-autosome fusion.** *Chromosome Res* 2007, **15:**223–230.
19. Veyrunes F, Chevret P, Catalan J, Castiglia R, Watson J, Dobigny G, Robinson TJ, Britton-Davidian J: **A novel sex determination system in a close relative of the house mouse.** *Proc R Soc B* 2010, **277:**1049–1056.

20. Lecompte E, ter Meulen J, Emonet S, Daffis S, Charrel RN: **Genetic identification of Kodoko virus, a novel arenavirus of the African pigmy mouse (*Mus Nannomys minutoides*) in West Africa.** *Virology* 2007, **364:**178–183.
21. de Bellocq JG, Borremans B, Katakweba A, Makundi R, Baird SJE, Becker-Ziaja B, Günther S, Leirs L: **Sympatric occurrence of 3 arenaviruses, Tanzania.** *Emerg Infect Dis* 2010, **16**(4):692–695.
22. Coulibaly-N'Golo D, Allali B, Kan Kouassi S, Fichet-Calvet E, Becker-Ziaja B, Rieger T, Ölschläger S, Dosso H, Denys C, ter Meulen J, Akoua-Koffi C, Günther S: **Novel arenavirus sequences in *Hylomyscus* sp. and *Mus* (*Nannomys*) *setulosus* from Cote d'Ivoire: Implications for evolution of arenaviruses in Africa.** *PLoS One* 2011, **6:**e20893.
23. Kronmann KC, Nimo-Paintsil S, Guirguis F, Kronmann LC, Bonney K, Obiri-Danso K, Ampofo W, Fichet-Calvet E: **Two novel arenaviruses detected in Pygmy mice, Ghana.** *Emerg Infect Dis* 2013, **19**(11):1832–1835.
24. Winkler AJ: **Neogene paleobiogeography and East African paleoenvironments: contributions from the Tugen Hills rodents and lagomorphs.** *J Hum Evol* 2002, **42:**237–256.
25. Lecompte E, Granjon L, Peterhans JK, Denys C: **Cytochrome b-based phylogeny of the *Praomys* group (Rodentia, Murinae): a new African radiation?** *C R Biol* 2002, **325**(7):827–840.
26. Stanhope MJ, Czelusniak J, Si JS, Nickeson J, Goodman M: **A molecular perspective on mammalian evolution from the gene encoding interphotoreceptor retinoid binding protein, with convincing evidence for bat monophyly.** *Mol Phylogenet Evol* 1992, **1:**148–160.
27. Galan M, Pagès M, Cosson JF: **Next-generation sequencing for rodent barcoding: species identification from fresh, degraded and environmental samples.** *PLoS One* 2012, **7**(11):e48374.
28. Stamatakis A: **RAxML Version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies.** *Bioinformatics* 2014, **30**(9):1312–1313.
29. Ronquist F, Huelsenbeck JP: **MrBayes 3: Bayesian phylogenetic inference under mixed models.** *Bioinformatics* 2003, **19:**1572–1574.
30. Rambaut A, Drummond AJ: *Tracer v1.5*. 2007 [http://beast.bio.ed.ac.uk/Tracer]
31. Librado P, Rozas J: **DnaSP v5: a software for comprehensive analysis of DNA polymorphism data.** *Bioinformatics* 2009, **25:**1451–1452.
32. Pons J, Barraclough TG, Gomez-Zurita J, Cardoso A, Duran DP, Hazell S, Kamoun S, Sumlin WD, Vogler AP: **Sequence-based species delimitation for the DNA taxonomy of undescribed insects.** *Syst Biol* 2006, **55**(4):595–609.
33. Powel JR: **Accounting for uncertainty in species delineation during the analysis of environmental DNA sequence data.** *Methods Ecol Evol* 2012, **3:**1–11.
34. Fujisawa T, Barraclough TG: **Delimiting species using single-locus data and the Generalized Mixed Yule Coalescent approach: A revised method and evaluation on simulated data sets.** *Syst Biol* 2013, **62**(5):707–724.
35. Drummond AJ, Suchard MA, Xie D, Rambaut A: **Bayesian phylogenetics with BEAUti and the BEAST 1.7.** *Mol Biol Evol* 2012, **29:**1969–1973.
36. Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S: **MEGA5: Molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods.** *Mol Biol Evol* 2011, **28:**2731–2739.
37. Baker RJ, Bradley RD: **Speciation in mammals and the Genetic Species Concept.** *J Mammal* 2006, **87:**643–662.
38. Lecompte E, Aplin K, Denys C, Catzeflis F, Chades M, Chevret P: **Phylogeny and biogeography of African Murinae based on mitochondrial and nuclear gene sequences with a new tribal classification of the subfamily.** *BMC Evol Biol* 2008, **8:**e199.
39. Schenk JJ, Rowe KC, Steppan SJ: **Ecological opportunity and incumbency in the diversification of repeated continental colonizations by muroid rodents.** *Syst Biol* 2013, **62:**837–864.
40. Jacobs LL, Flynn LJ: **Of mice… again: the Siwalik rodent record, murine distribution, and molecular clocks**. In *Interpreting the past: essays on human, primate and mammal evolution.* Edited by Lieberman D, Smith R, Kelley J. Leiden: Brill Academic Publishers; 2005:63–80.
41. Denys C: **Deux nouvelles espèces d'*Aethomys* (Rodentia, Muridae) à Langebaanweg (Pliocène, Afrique du Sud): Implications phylogénétiques.** *Annales de Paléontologie* 1990, **76:**41–69.
42. Ambrose SH, Bell CJ, Bernor RL, Boisserie J-R, Darwent CM, Degusta D, Deino A, Garcia N, Haile-Selassie Y, Head JJ, Howell FC, Kyule MD, Manthi FK, Mathu EM, Nyamai CM, Saegusa H, Stidham TA, Williams MAJ, Hlusko LJ: **The paleoecology and paleogeographic context of Lemudong'o locality 1, a Late Miocene terrestrial fossil site in southern Kenya.** *Kirtlandia* 2007, **56:**38–52.
43. Mein P, Pickford M, Senut B: **Late Miocene micromammals from the Harasib karst deposits, Namibia: Part 2b. Cricetomyidae, Dendromuridae and Muridae, with an addendum on the Myocricetodontinae.** *Commun Geological Surv Namibia* 2004, **13:**43–61.
44. Fabre PH, Hautier L, Dimitrov D, Douzery EJP: **A glimpse on the pattern of rodent diversification: a phylogenetic approach.** *BMC Evol Biol* 2012, **12:**e88.
45. Lemey P, Rambaut A, Drummond AJ, Suchard MA: **Bayesian phylogeography finds its roots.** *PLoS Comput Biol* 2009, **5:**e1000520.
46. Ree RH, Smith SA: **Maximum likelihood inference of geographic range evolution by dispersal, local extinction, and cladogenesis.** *Syst Biol* 2008, **57**(1):4–14.
47. Ree RH, Moore BR, Webb CO, Donoghue MJ: **A likelihood framework for inferring the evolution of geographic range in phylogenetic trees.** *Evolution* 2005, **59**(11):2299–2311.
48. R Core Team: *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing; 2013 [http://www.R-project.org/]
49. Galtier N, Gouy M, Gautier C: **SEAVIEW and PHYLO_WIN: two graphic tools for sequence alignment and molecular phylogeny.** *Comput Appl Biosci* 1996, **12:**543–548.
50. Yalden DW: **Small mammals of the Bale Mountains, Ethiopia.** *Afr J Ecol* 1988, **26:**281–294.
51. Lavrenchenko LA: **The mammals of the isolated Harenna Forest (southern Ethiopia): structure and history of the fauna**. In *Isolated Vertebrate Communities in the Tropics*, Proc. 4th Int. Symp. Bonn, Bonn Zool Monogr. Edited by Rheinwald G. ; 2000. **46:**223-231.
52. Aniskin VM, Lavrenchenko LA, Varshavskii AA, Milishnikov AN: **Karyotypes and cytogenetic differentiation of two African mouse species of genus *Mus* (Rodentia, Muridae).** *Rus J Genet* 1998, **34**(1):80–85.
53. Yalden DW, Largen MJ, Kock D, Hillman JC: **Catalogue of the mammals of Ethiopia and Eritrea. 7. Revised checklist, zoogeography and conservation.** *Trop Zool* 1996, **9:**73–164.
54. Dieterlen F, Happold DCD: ***Mus triton* Grey-bellied Pygmy Mouse.** In *Mammals of Africa, Volume III Rodents, hares and rabbits.* Edited by Happold DCD. London and New York: Bloomsbury Publishing; 2013:497–499.
55. Yalden DW, Largen MJ: **The endemic mammals of Ethiopia.** *Mamm Rev* 1992, **22:**115–150.
56. Vermeiren L, Verheyen WN: **Notes sur les *Leggada* de Lamto, Cote d'Ivoire, avec la description de *Leggada baoulei* sp. n. (Mammalia, Muridae).** *Revue de Zoologie Africaine* 1980, **94:**570–590.
57. Petter F: **Les souris africaines du groupe *sorella* (Rongeurs, Muridés).** *Mammalia* 1981, **45**(3):313–320.
58. Mboumba JF, Deleporte P, Colyn M, Nicolas V: **Phylogeography of *Mus* (*Nannomys*) *minutoides* (Rodentia, Muridae) in West Central African savannahs: singular vicariance in neighbouring populations.** *J Zool Syst Evol Res* 2011, **49**(1):77–85.
59. Brouat C, Tatard C, Bâ K, Cosson JF, Dobigny G, Fichet-Calvet E, Granjon L, Lecompte E, Loiseau A, Mouline K, Duplantier JM: **Phylogeography of the Guinea Multimammate Mouse (*Mastomys erythroleucus*): a case study for Sahelian species in West Africa.** *J Biogeogr* 2009, **36:**2237–2250.
60. Bryja J, Patzenhauerová H, Granjon L, Dobigny G, Konečný A, Duplantier JM, Gauthier P, Colyn M, Durnez L, Lalis A, Nicolas V: **Plio-Pleistocene history of West African Sudanian savanna and the phylogeography of the *Praomys daltoni* complex (Rodentia): the environment/geography/genetic interplay.** *Mol Ecol* 2010, **19:**4783–4799.
61. Dobigny G, Tatard C, Gauthier P, Ba K, Duplantier JM, Granjon L, Kergoat GJ: **Mitochondrial and nuclear genes-based phylogeography of *Arvicanthis niloticus* (Murinae) and sub-Saharan open habitats Pleistocene history.** *PLoS One* 2013, **8**(11):e77815.
62. Colangelo P, Corti M, Verheyen E, Annesi F, Oguge N, Makundi RH, Verheyen W: **Mitochondrial phylogeny reveals differential modes of chromosomal evolution in the genus *Tatera* (Rodentia: Gerbillinae) in Africa.** *Mol Phylogenet Evol* 2005, **35:**556–568.
63. McDonough MM, Sotero-Caio CG, Ferguson AW, Lewis PJ, Tswiio M, Thies ML: **Mitochondrial DNA and karyotypic data confirm the presence of *Mus indutus* and *Mus minutoides* (Mammalia, Rodentia, Muridae, *Nannomys*) in Botswana.** *Zookeys* 2013, **359:**35–51.
64. Chevret P, Robinson TJ, Perez J, Veyrunes F, Britton-Davidian J: **A phylogeographic survey of the Pygmy mouse *Mus minutoides* in South Africa: Taxonomic and karyotypic inference from cytochrome *b* sequences of museum specimens.** *PLoS One* 2014, **9**(6):e98499.

65. Dobigny G, Tatard C, Kane M, Gauthier P, Brouat C, Ba K, Duplantier JM: **A cytotaxonomic and DNA-based survey of rodents from Northern Cameroon and Western Chad.** *Mammal Biol* 2011, **76**(4):417–427.
66. Winkler AJ, Denys C, Avery MD: **Rodentia**. In *Cenozoic Mammals of Africa.* Edited by Werdelin L, Sanders WJ. Berkeley: University of California Press; 2010:263–304.
67. Isaac JB, Mallet J, Mace GM: **Taxonomic inflation: its influence on macroecology and conservation.** *Trends Ecol Evol* 2004, **19:**464–469.
68. Lavrenchenko LA, Verheyen WN, Verheyen E, Hulselmans J, Leirs H: **Morphometric and genetic study of Ethiopian *Lophuromys flavopunctatus* Thomas, 1888 species complex with description of three new 70-chromosomal species (Muridae, Rodentia).** *Bulletin de l'Insitut Royal des Sciences Naturelles de Belgique, Biologie* 2007, **77:**77–117.
69. Berger SA, Krompass D, Stamatakis A: **Performance, accuracy, and web server for evolutionary placement of short sequence reads under maximum likelihood.** *Syst Biol* 2011, **60:**291–302.
70. Skinner JD, Chimimba CT: *The Mammals of the Southern African Subregion.* 3rd edition. Cambridge, UK: Cambridge University Press; 2005.
71. Monadjem A: ***Mus setzeri* Setzer's Pygmy Mouse**. In *Mammals of Africa, Volume III Rodents, hares and rabbits.* Edited by Happold DCD. London and New York: Bloomsbury Publishing; 2013:493–494.
72. Vermeiren L, Verheyen WN: **Additional data on *Mus setzeri* Petter (Mammalia, Muridae).** *Annales Musée Royal de l'Afrique Centrale, ser 8 (Sciences Zoologiques)* 1983, **237:**137–141.
73. Missoup AD, Nicolas V, Wendelen W, Keming E, Bilong Bilong CF, Couloux A, Atanga E, Hutterer R, Denys C: **Systematics and diversification of *Praomys* species (Rodentia: Muridae) endemic to the Cameroon Volcanic Line (West Central Africa).** *Zool Scr* 2012, **41:**327–345.
74. Demos TC, Agwanda B, Hickerson MJ: **Integrative taxonomy within the *Hylomyscus denniae* complex (Rodentia: Muridae) and a new species from Kenya.** *J Mamm* 2014, **95**(1):1–15.
75. Verheyen WN, Hulselmans JLJ, Dierckx T, Mulungu L, Verheyen E, Corti M, Kerbis JP, Leirs H: **Craniometric,cytogenetic and genetic characterization of the Kilimanjaro *Lophuromys aquilus* True 1892 population, its implications for the taxonomy of the *Lophuromys flavopunctatus* species complex and description of five new taxa. (Rodentia-Muridae-Africa).** *Bulletin de l'Insitut Royal des Sciences Naturelles de Belgique, Biologie* 2007, **77:**23–75.
76. Colangelo P, Verheyen E, Leirs H, Tatard C, Denys C, Dobigny G, Duplantier JM, Brouat C, Granjon L, Lecompte E: **A mitochondrial phylogeographic scenario for the most widespread African rodent, *Mastomys natalensis*.** *Biol J Linn Soc* 2013, **108**(4):901–916.
77. Potts R: **Hominin evolution in settings of strong environmental variability.** *Quat Sci Rev* 2013, **73:**1–13.
78. Alvarado-Serrano DF, Knowles LL: **Ecological niche models in phylogeographic studies: applications, advances and precautions.** *Mol Ecol Resour* 2014, **14:**233–248.
79. Šmíd J, Carranza S, Kratochvíl L, Gvoždík V, Nasher AK, Moravec J: **Out of Arabia: A complex biogeographic history of multiple vicariance and dispersal events in the gecko genus *Hemidactylus* (Reptilia: Gekkonidae).** *PLoS One* 2013, **8**(5):e64018.
80. Portik DM, Papenfuss TJ: **Monitors cross the Red Sea: The biogeographic history of *Varanus yemenensis*.** *Mol Phylogenet Evol* 2012, **62:**561–565.
81. Winney BJ, Hammond RL, Macasero W, Flores B, Boug A, Biquand V, Biquand S, Bruford MW: **Crossing the Red Sea: phylogeography of the hamadryas baboon, *Papio hamadryas hamadryas*.** *Mol Ecol* 2004, **13:**2819–2927.
82. Bosworth W, Huchon P, McClay K: **The red sea and gulf of aden basins.** *J Afr Earth Sci* 2005, **43:**334–378.
83. Fernandes CA, Rohling EJ, Siddall M: **Absence of post-Miocene Red Sea land bridges: biogeographic implications.** *J Biogeogr* 2006, **33:**961–966.
84. Catzeflis FM, Denys C: **The African *Nannomys* (Muridae): an early offshoot from the *Mus* lineage - Evidence from scnDNA hybridization experiments and compared morphology.** *Isr J Zool* 1992, **38:**219–231.
85. Denys C: **Of mice and men: Evolution in East and South Africa during Plio-Pleistocene times**. In *African Biogeography, Climate Change, and Human Evolution.* Edited by Bromage TG, Schrenk F. Oxford: Oxford University Press; 1999:226–252.
86. Trauth MH, Maslin MA, Deino AL, Strecker MR, Bergner AGN, Dühnforth M: **High- and low-latitude forcing of Plio-Pleistocene East African climate and human evolution.** *J Hum Evol* 2007, **53:**475–486.
87. de Menocal PB: **African climate change and faunal evolution during the Pliocene-Pleistocene.** *Earth Planet Sc Lett* 2004, **220:**3–24.
88. Zachos J, Pagani M, Sloan L, Thomas E, Billups K: **Trends, rhythms, and abberations in global climate 65 Ma to present.** *Science* 2001, **292:**686–693.
89. Schultz S, Maslin M: **Early human speciation, brain expansion and dispersal influenced by African climate pulses.** *PLoS One* 2013, **8**(10):e76750.
90. Veyrunes F, Catalan J, Tatard C, Cellier-Holzem E, Watson J, Chevret P, Robinson TJ, Britton-Davidian J: **Mitochondrial and chromosomal insights into karyotypic evolution of the pygmy mouse, *Mus minutoides*, in South Africa.** *Chromosome Res* 2010, **18:**563–574.
91. Veyrunes F, Perez J, Paintsil SNC, Fichet-Calvet E, Britton-Davidian J: **Insights into the evolutionary history of the X-linked sex reversal mutation in *Mus minutoides*: clues from sequence analyses of the Y-linked Sry gene.** *Sex Dev* 2013, **7**(5):244–252.
92. Matthey R: **Nouvelles contributions à la cytogénétique des Mus africaines du sous-genre *Leggada*.** *Experientia* 1966, **22**(6):400–401.
93. Nichol ST, Arikawa J, Kawaoka Y: **Emerging viral diseases.** *Proc Natl Acad Sci U S A* 2000, **2000**(97):12411–12412.
94. Günther S, Lenz O: **Lassa virus.** *Crit Rev Clin Lab Sci* 2004, **41:**339–390.
95. Hugot JP, Gonzales JP, Denys C: **Evolution of the Old World Arenaviridae and their rodent hosts: generalized host-transfer or association by descent?** *Infect Genet Evol* 2001, **1:**13–20.

Chapter IV

Revised occurrence of rodents from the tribe Praomyini (Muridae) in Zambia based on mitochondrial DNA analyses: implications for biogeography and conservation.

Bryja J., **Mazoch V.**, Patzenhauerová H., Mateke C., Zima J. Jr, Šklíba J. & Šumbera R. (2012), (2012). *Folia Zoologica*, 61(3/4), 268-283.

Folia Zool. – 61 (3–4): 268–283 (2012)

Revised occurrence of rodents from the tribe Praomyini (Muridae) in Zambia based on mitochondrial DNA analyses: implications for biogeography and conservation

Josef BRYJA[1]*, Vladimír MAZOCH[2], Hana PATZENHAUEROVÁ[1], Clare MATEKE[3], Jan ZIMA jr.[2], Jan ŠKLÍBA[2] and Radim ŠUMBERA[2]

[1] *Institute of Vertebrate Biology, Academy of Sciences of the Czech Republic, v.v.i., Květná 8, 603 65 Brno, Czech Republic; e-mail: bryja@brno.cas.cz*
[2] *Department of Zoology, Faculty of Science, University of South Bohemia, Branišovská 31, 370 05 České Budějovice, Czech Republic*
[3] *Livingstone Museum, Mosi-o-Tunya Road, PO Box 60498, Livingstone, Zambia*

Received 19 January 2012; Accepted 25 June 2012

Abstract. The taxonomy and distribution of rodents in Zambia was comprehensively summarized in 1978 by W.F.H. Ansell in his excellent book Mammals of Zambia. Despite the fact that during the last three decades many new taxonomic revisions of African rodents were published and extensive new material collected, not much work has been done on Zambian rodents since the book publication. Here we summarize the current knowledge of one of the most speciose group of African rodents, the tribe Praomyini, in Zambia. We review available historical records and revise our recently collected material by sequencing the mitochondrial DNA gene of cytochrome *b*. The presence of eight species of Praomyini in Zambia is documented and the pattern of their geographical distribution is described and discussed. Two species, *Praomys minor* and *Mastomys coucha*, are reported for the first time from Zambia and *Praomys* cf. *jacksoni* probably represents a new undescribed species. On the other hand, the actual occurrence of *Colomys goslingi*, known in Zambia only from one historical record, is questionable. The results document the usefulness of the DNA barcoding approach for description of species diversity of taxonomically complicated groups with many cryptic species.

Key words: zoogeography, faunistics, mtDNA, Murinae, phylogeny, Rodentia, DNA barcoding

Introduction

Rodents are a very important group of mammals forming more than 42 % of mammalian species diversity (Musser & Carleton 2005). They have a practical importance for humans, being suitable models for research e.g. in biomedicine (e.g. laboratory mouse or rat) or evolutionary ecology (e.g. some voles as models in population demography). They are also significant pests in agriculture and vectors of many infectious diseases (e.g. Gratz 2006, Singleton et al. 2010). Besides intensive research of rodents across the globe over the last century, there are still many geographical areas, where even basic data are missing and the rodent diversity is poorly described. Filling the gap in this knowledge has many important consequences, e.g. for proposing new areas for biodiversity conservation or for epidemiological studies.

Among the least known areas in this sense, some parts of tropical Africa belong to the most intriguing for many reasons. During last five millions years many groups of African rodents underwent a rapid radiations linked with the speciation processes, resulting in extreme cryptic diversity. The use of DNA-based methods led to the discovery of many monophyletic genetic lineages within phenotypically

* *Corresponding Author*

uniform groups that were subsequently described as separate biological species (e.g. Hoffmann et al. 2009). This approach is very useful, because it allows quick identification of the centres of genetic diversity even in the lack of classical taxonomical knowledge (e.g. Kan Kouassi et al. 2008, Dobigny et al. 2011).

The Zambezian phytochorion, and Zambia particularly, is an important geographical area, where the diversity of mammalian fauna has been studied only marginally by molecular-genetic approaches. Rodents of Zambia were intensively studied up to the 1970s of the last century and the knowledge of their taxonomy and distribution are summarized in an excellent book, The Mammals of Zambia by W.F.H. Ansell (Ansell 1978). Since that time, however, the rodents of Zambia were studied only occasionally as part of geographically localized ecological studies (e.g. Chidumayo 1977, 1979, 1980a, b, 1984, Kawalika 2004) and a few individuals from Zambia were also used as a part of recent systematic reviews of some taxonomic groups (e.g. Burda et al. 1999, Castiglia et al. 2002, Verheyen et al. 2003, Castiglia et al. 2003, Corti et al. 2004, 2005, Mullin et al. 2004, 2005, Carleton & Stanley 2005, Taylor et al. 2009, Verheyen et al. 2011). The lack of systematic research of Zambian rodents in the last 40 years has led to the absence of up-to-date patterns of distribution of rodent species, especially in the taxonomically complicated groups, where the genetic approaches recently led to the description of many new cryptic species (e.g. Van Daele et al. 2007). Absence of recent systematic research in this area is quite surprising, because the Zambezian Region is the largest major phytochorion in Africa after the Sahara (White 1983). Zambia itself is a large country and despite the fact that the most landscape is covered with relatively uniform miombo woodland, there is still a wide spectrum of other habitats suggesting higher diversity of small mammals than expected so far. In lower altitudes, dominant miombo is substituted with mopane woodland, and there are also other types of savannah woodlands and shrubs across the country. Several types of forests such as evergreen forests especially in northern Zambia, gallery forest along rivers and patches of Afromontane evergreen forest in the eastern highlands together with seasonal floodplains, dambos and wetlands, grasslands and thickets contribute to a high diversity of habitats suitable for small mammals.

The tribe Praomyini (sensu Lecompte et al. 2008) belongs to the family Muridae (subfamily Murinae) and it is one of the most intensively studied groups of African rodents. This endemic African taxon is interesting and important for several reasons. It comprises one of the most important agricultural pests (e.g. in the genus *Mastomys*), and their population ecology, demography and direct impact in agriculture have been intensively studied (e.g. Granjon et al. 2005, Sluydts et al. 2009). They are also hosts of many pathogens and some of them are very important for human health (e.g. Lassa virus in western Africa; Lecompte et al. 2006). A recent description of the highly pathogenic virus Lujo (Briese et al. 2009) led to intensive searches of new pathogens, which resulted in discoveries of new arenaviruses (Coulibaly-N'Golo et al. 2011, Ishii et al. 2011), or polyomaviruses (Orba et al. 2011) hosted by Praomyini rodents. Despite this importance for human being, the basic knowledge of many taxa is still very poor. Several genera of the tribe were recently revised using a combination of DNA-based methods and morphology and candidates for new species were identified (e.g. Dobigny et al. 2008, Nicolas et al. 2008a, 2010, Bryja et al. 2010, Kennis et al. 2011). More detailed information about the distribution and ecology of these speciose genera is strongly required, for example, for the improved understanding of the epidemics and dynamics of emerging infectious diseases (Fichet-Calvet et al. 2007, 2008, Coulibaly-N'Golo et al. 2011).

In this study, we review the occurrence of species of the Praomyini tribe in Zambia. Based on the recent definition of the tribe (Lecompte et al. 2008) and faunistic data of Ansell (1978), the Praomyini is in Zambia represented by six species in the genera *Mastomys*, *Praomys*, *Hylomyscus*, *Zelotomys* and *Colomys*. However, the first three genera underwent important taxonomic revisions during last decade, which can obscure previous faunistic data. Within our recent projects, two field expeditions to Zambia were organized in 2009-2010 and numerous material of rodents from diverse habitats at many localities across the whole country were collected. Here we use the sequences of mitochondrial DNA and phylogenetic methods to assign individuals into species of recently revised genera. Together with previously published data, we present the current status of knowledge on distribution of these interesting rodents in Zambia and we highlight the prospective directions of future research of this group.

Material and Methods

Geography of Zambia

Zambia is a land-locked country covering 752620 km^2 in south-central Africa, between 10 and 18 degrees south of the Equator. It is bordered to the east by

Malawi and Tanzania, north by the Democratic Republic of the Congo (DRC), west by Angola and to the south by Namibia, Botswana, Zimbabwe and Mozambique (Jachmann 2000) (Fig. 1). The country is largely occupied by the undulating Central African Plateau, varying from about 900 to 1400 m above sea level. At some places, particularly at the northern-eastern rim it rises to over 2000 m, reaching 2320 m on the Mafingas, whereas it drops down to 329 m in the Zambezi Valley. At the east, the Plateau is delimited by the Great Escarpment, the southern extension of the Great Rift Valley. The southern border of Zambia is made by the River Zambezi, on which the famous Victoria Falls (Mosi-o-Tunya) is located.

Zambia has a mild tropical climate with three distinct seasons: the cool dry season from May to August, the hot dry season from September to October and the warm wet season from November to April. Maximum daytime temperatures in the hot season average 27-38 °C, with the highest temperatures in the south-west and in the rift valley areas, and the lowest temperatures in the north-east on the high plateau; while minimum daily cold-season temperatures average between 2-15 °C, with the lowest temperatures in the south-west (Jachmann 2000). Mean annual rainfall is between 600 mm in the South and 1600 mm in the North (Copperbelt, Luapula and Northern Province) (Fig. 1b). Zambia is drained by two major river basins. The Zambezi River Basin in the South (about 3/4 of the country's total area) consists of the Zambezi and its major tributaries (the Kafue and the Luangwa). The Congo Basin in the North has two major rivers, Chambeshi and Luapula. From the point of view of phytochorial classification, the whole of Zambia belongs (together with Malawi, Zimbabwe, large parts of Angola, Tanzania and Mozambique) to the so-called Zambezian Region, represented by nine ecoregions in four biomes. The most widespread are Miombo, Mopane and Baikiaea woodland savannah, with grasslands (mainly flooded) and small patches of evergreen forest (see White 1983; Mayaux et al. 2004, and Fig. 1a for more detailed information on vegetation types).

Sampling

Two field expeditions to Zambia in 2009 and 2010 allowed visiting a total of 57 localities (Fig. 1). Small mammals were captured along transect lines during 70 trapping nights using wooden snap traps and medium-sized (8 × 9 × 23 cm) Sherman live traps that were set on the ground 5-15 m apart from each other depending on habitat type. Trapping effort per locality varied significantly, but typically cca 50-200 traps (combination of different types) were installed for 1-2 nights. Captured animals were examined and provisionally identified to genera based on external morphology, measurements and visible external features. At least one individual from each genus per locality was used for genetic analysis. In most localities, however, we sequenced more individuals (see Supplementary Table S1) and we especially tried to analyse morphologically different animals (e.g. with shortest and longest tail). Preferably spleen or kidney tissues were collected from sacrificed individuals and toe clips from subsequently released live animals. Tissue samples were preserved in the 96 % ethanol and stored at –20 °C until DNA extraction. Voucher specimens are deposited in the Vertebrate Collection of the University of South Bohemia (České Budějovice, Czech Republic), mammalian collection of the Livingstone museum (Livingstone, Zambia) and in the Copperbelt Museum (Ndola, Zambia).

DNA extraction, PCR and phylogenetic analyses

The DNA was extracted using the DNeasy Blood & Tissue Kit (Qiagen) following the manufacturer's instructions. The complete cytochrome *b* gene was amplified by polymerase chain reaction (PCR) using the primers L14723 and H15915 (Lecompte et al. 2002). We also extracted DNA from tissues (toes) of selected samples from the Livingstone Museum using the same extraction kit. Extracted DNA was however very fragmented and we had to use a set of additional internal primers H15553, L15408, L15146, H15149 and L15513 (Lecompte et al. 2002) to obtain a sequence of cytochrome *b*. The PCR conditions were the same for all primer pairs – the cycling procedure started at 94 °C of initial denaturation for 3 min, followed by 35 cycles of 94 °C (30 sec), 50 °C (30 sec) and 72 °C (3 min), ending with final extension at 72 °C for 10 min. Sequencing of obtained amplicons was performed using BigDyeTM terminator chemistry (Applied Biosystems). All sequences are available in GenBank under accession numbers JX126523-JX126621 (Supplementary Table S1).

Sequence electrophoretograms were edited in SeqScape v. 2.5.0 (Applied Biosystems) and subsequently aligned with available GenBank sequences for particular genera (representing at least one sequence per intraspecific clade identified in previous studies; see Supplementary Table S1 and below for more details). These datasets were used for phylogenetic analysis using a maximum likelihood (ML) method with the aim to identify the species of

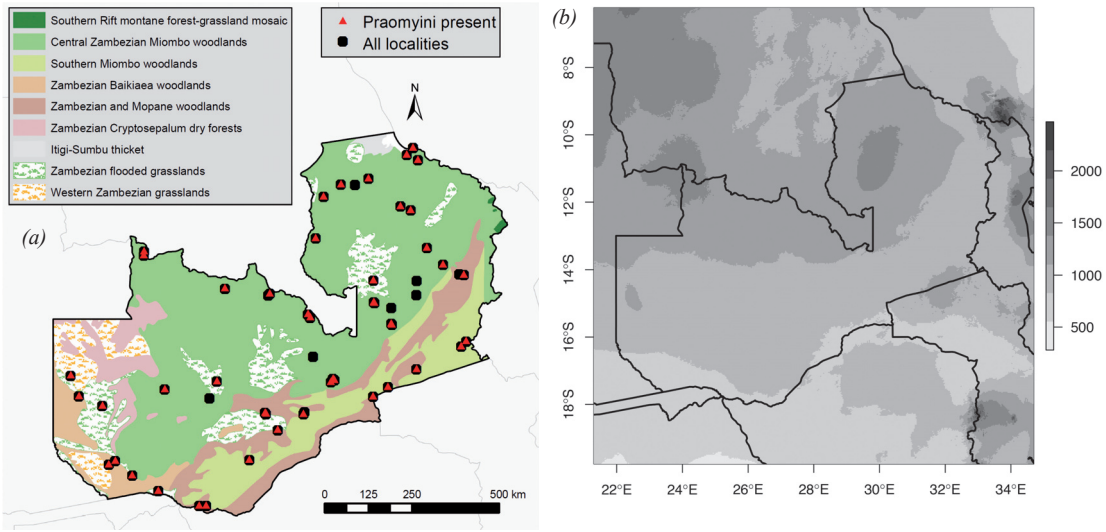


Fig. 1. *(a) Map of terrestrial ecoregions of Zambia (modified from Olson et al. 2001) with a distribution of trapping sites (black squares) and localities with the presence of species of the Praomyini tribe (red triangles) during our field expeditions. (b) Map of mean annual rainfall in Zambia.*

Zambian individuals. The online program FindModel (http://www.hiv.lanl.gov/content/sequence/findmodel/findmodel.html) was used to evaluate the fit of 28 nested models of nucleotide substitution to the data and the best model for each alignment was selected on the basis of Akaike information criterion (AIC). ML analyses were performed using PHYML online web server (Guindon et al. 2005) with the NNI algorithm and BIONJ distance-based tree as the starting tree. Bootstrap analysis (1000 replicates) was used to estimate the robustness of internal nodes. The results were visualised in FigTree v. 1.3.1 (http://tree.bio.ed.ac.uk/software/figtree/).

Analysis of published records and unpublished data from museum collections

Ansell (1978) revised the occurrence of all Zambian rodents up to 1978. We resurrected the GPS coordinates from the species distribution maps published in Ansell's book, by taking the centre of the particular quadrate as the position of the species occurrence. Furthermore, the data from the Livingstone museum (Livingstone, Zambia), African Rodentia (http://projects.biodiversity.be/africanrodentia, Access date: 20/12/2011) and GBIF (Vertebrate specimens, http://data.gbif.org/datasets/resource/541 and MVZ Mammal Catalog, http://data.gbif.org/datasets/resource/8121) databases collected after 1978 and additional data published after 1978 (Kawalika 2004) were reviewed and used for creating distribution maps using the software R 2.14.0 (R Development Core Team 2011) and *mapdata* package.

Results and Discussion

In total, we sampled rodents at 57 localities in Zambia and species from the tribe Praomyini occurred at 47 of them (Fig. 1; see list of localities and sequenced individuals in Supplementary Table S1). Below we describe the results of genetic analysis and annotate the current list of all known Praomyini species in Zambia.

(a) Genus Mastomys

The genus *Mastomys* belongs to the most intensively studied rodents in Africa (Granjon et al. 1997, Lecompte et al. 2005). It harbours widely distributed species (like *M. natalensis* which is known from most of sub-Saharan Africa) as well as much more localized species (e.g. *M. awashensis* known only from a few localities in Ethiopia; Colangelo et al. 2010). Most species avoid continuous tropical forests and some of them are known to be synanthropic. Ansell (1978) reported one species from Zambia, however, sequencing of mtDNA confirmed the presence of a second species.

Mastomys natalensis (Smith, 1834)

Natal multimammate mouse was the most frequently captured animal during our fieldwork and it is distributed throughout the whole country (Fig. 2a,

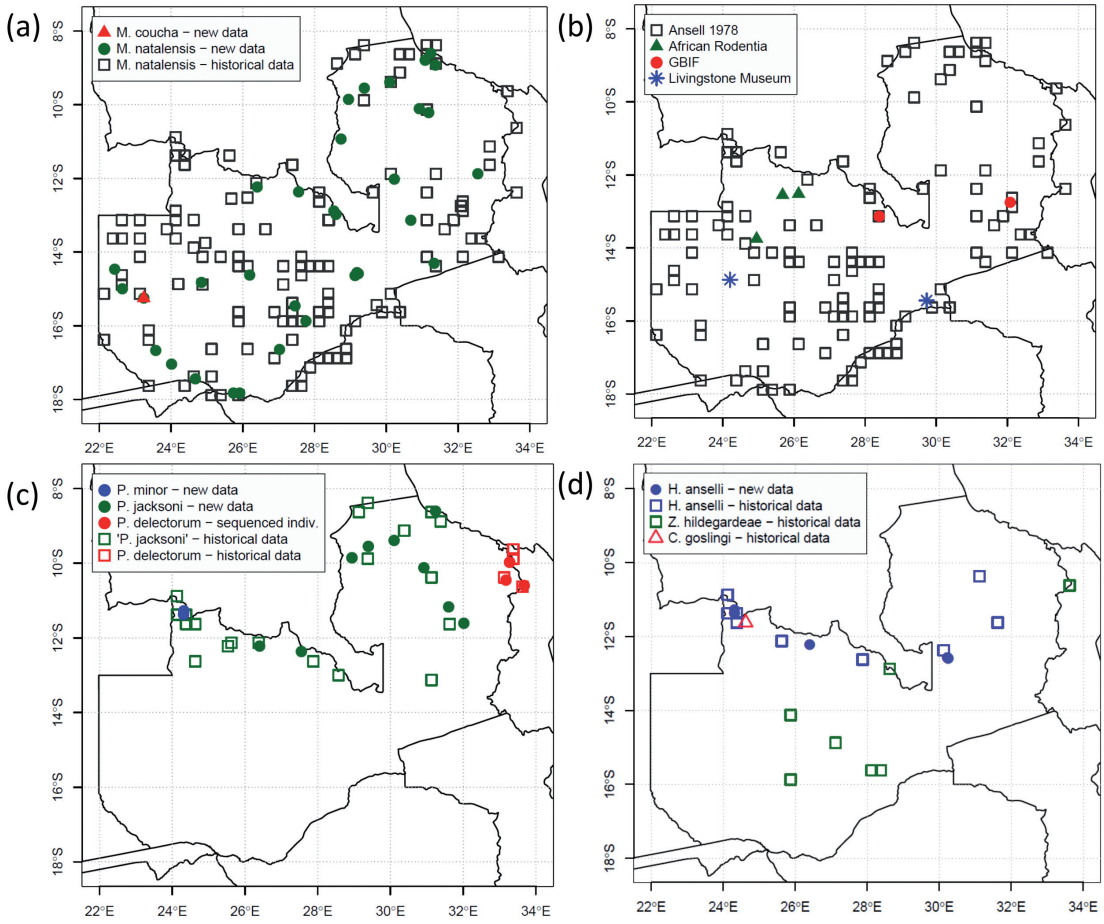


Fig. 2. *Distribution of eight species of Praomyini in Zambia. (a) genus* Mastomys*; (b) overview of historical records of* M. natalensis*; (c) genus* Praomys*; (d) genera* Hylomyscus*,* Zelotomys *and* Colomys*. In general, except part (b), filled symbols characterize the localities, where the material for our recent genetic analyses come from, while open symbols represent historical data ressurected from Ansell 1978 (+ the only more recent publication from Zambia, i.e. Kawalika 2004) and records from museum databases – Livingstone Museum (Livingstone, Zambia), African Rodentia (http://projects.biodiversity.be/africanrodentia, Access date: 20/12/2011) and GBIF data portal (Vertebrate specimens, http://data.gbif.org/datasets/resource/541 and MVZ Mammal Catalog, http://data.gbif.org/datasets/resource/8121).*

b). Phylogenetic analysis (Fig. 3a) clustered all Zambian individuals into a single cluster (but without bootstrap support) together with previously published sequences from Zambia and Botswana. Even usually strong barriers to gene flow, like Zambezi and Kafue Rivers, do not seem to have impact on the spatial structure of *M. natalensis* populations. However, we used only mtDNA sequences and more variable nuclear markers (like microsatellites) are necessary to identify the factors influencing genetic structure of this widely distributed rodent species.

Mastomys coucha (Smith, 1834)

This species is largely endemic to South African subregion. It is found in South Africa, Lesotho, southern and western Zimbabwe, Mozambique, Botswana, central and northern Namibia to southern Angola (Coetzee et al. 2008). Phylogenetic analysis clearly identified mtDNA sequences of three Zambian individuals from the close proximity of the Newa village near Mongu as *M. coucha* (Fig. 3a), which constitutes the first record for Zambia and the northernmost record of this species (Supplementary

Table S1). The species was sympatric and even syntopic with *M. natalensis* in wet fallow fields near a stream.

(b) Genus Praomys

The species-rich genus *Praomys* consists of predominantly forest species widely distributed in tropical Africa. It was recently reviewed using both morphological and molecular data, one new species was described (Van der Straeten 2008) and several other candidates for the description of new species were identified (e.g. Akpatou et al. 2007, Bryja et al. 2010, Nicolas et al. 2005, 2008b, 2010, 2011). However, none of recent studies included material from the Zambezian phytochorion. Our molecular data provide clear evidence of the occurrence of three species in Zambia (Fig. 2c), one of them is reported for the first time from the country.

Praomys cf. *jacksoni* (de Winton, 1897)

"*Praomys jacksoni*" was reported from 18 quadrates by Ansell (1978), but the results of molecular analysis indicate the presence of two cryptic species in Zambia, *P.* cf. *jacksoni* and *P. minor* (Fig. 3b). *P.* cf. *jacksoni* is the most widespread species of *Praomys* in Zambia and the majority of previous records of "*P. jacksoni*" (sensu Ansell 1978) probably represent this species. It is distributed in the northern part of the country, i.e. in areas with more than 1100 mm annual rainfall (Fig. 1). All sequenced *P.* cf. *jacksoni* from Zambia form a monophyletic group that is a sister clade to *P.* cf. *jacksoni* Clade II (sensu Kennis et al. 2011) from the Kisangani region in DRC (Fig. 3b). The mean K2P distance between these two clades is 3.1 ± 0.63. The individuals from DRC are distributed on the right bank of the River Congo and they morphologically differ from nominate species (*P. jacksoni* s. str.) described from Uganda (see Fig. 4 in Kennis et al. 2011). The whole group (i.e. Clade II from DRC together with individuals from Zambia) could therefore represent a new species; however more material (especially from the eastern part of the Albertine rift), genetic data (nuclear genes) and morphological analyses (including e.g. geometric morphometric approach) are required for final decision about its specific status.

Praomys minor Hatt, 1934

All 18 sequenced individuals from two localities in the Mwinilunga District (Supplementary Table S1) in the north-western part of Zambia (where the River Zambezi springs) cluster to a separate group forming a monophyletic lineage together with Clade III from DRC (sensu Kennis et al. 2011) (Fig. 3b). The mean K2P distance between Zambian and DRC clades is 3.1 ± 0.59. The individuals from DRC occur at the left bank of the River Congo in the Kisangani region and are morphologically distinct from other clades of the *P. jacksoni* complex (see Fig. 4 in Kennis et al. 2011), especially they are significantly smaller at all skull measurements than other taxa from the complex (see Table 3 in Kennis et al. 2011). They are morphologically very similar to type series of *P. minor*, i.e. the species described on the basis of three specimens from Lukolela in DRC by Hatt (1934). The material published by Kennis et al. (2011) is thus very probably the first well-documented record of *P. minor* after its original description extending the geographical range of the species about 750 km to the east. Here, we extend its distribution range by more than 1300 km to the south. It is a new species for Zambia, where it reaches the southern border of its distribution. The species was very abundant at both localities in Zambia (gallery forest and moist evergreen forest called "mushitu") and occupied various habitats in Kisangani region in DRC (Kennis et al. 2011). This suggests that it was largely overlooked in previous studies of rodents in the Congo Basin, where it is probably largely distributed on the left bank of the River Congo. We have no evidence that *P. minor* at Mwinilunga District is sympatric with *P.* cf. *jacksoni*, suggesting that all (or most) of *P. jacksoni* sensu Ansell (1978) from this area are probably *P. minor*. However, more sequenced individuals from more localities are necessary to obtain a more precise record of the distribution of both species in north-western Zambia.

Praomys delectorum (Thomas, 1910)

This mountain species has been recorded from high plateaus and isolated mountains from northern Mozambique, through Malawi, north-eastern Zambia and Tanzania (Eastern Arc Mountains, Kilimanjaro, Mt. Meru) to south-eastern Kenya (Taita Hills) (Bryja et al., submitted). In Zambia, the species is abundant in the forest patches on Nyika Plateau and surrounding mountains (the Makutus and the Mafingas) (Ansell 1978; Fig. 2c). Recent detailed morphological study suggested that individuals from Zambia, northern Malawi and southern Tanzania should belong to a separate species *P. melanotus* Allen & Loveridge, 1933 (Carleton & Stanley 2012), however, in this study we still consider it as *P. delectorum*. Within the larger phylogeographic study of the species (Bryja et al., submitted), we sequenced historical DNA of five

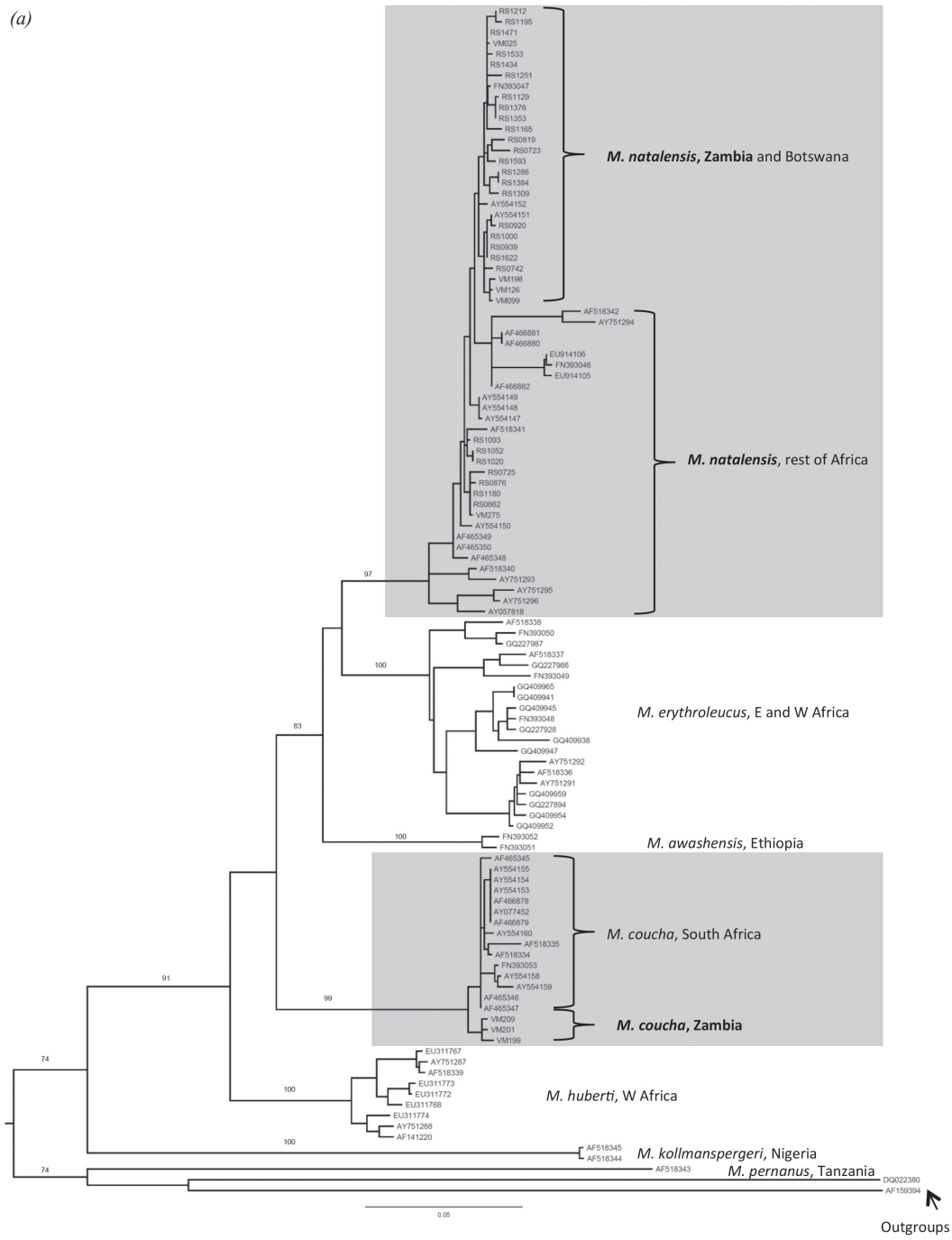
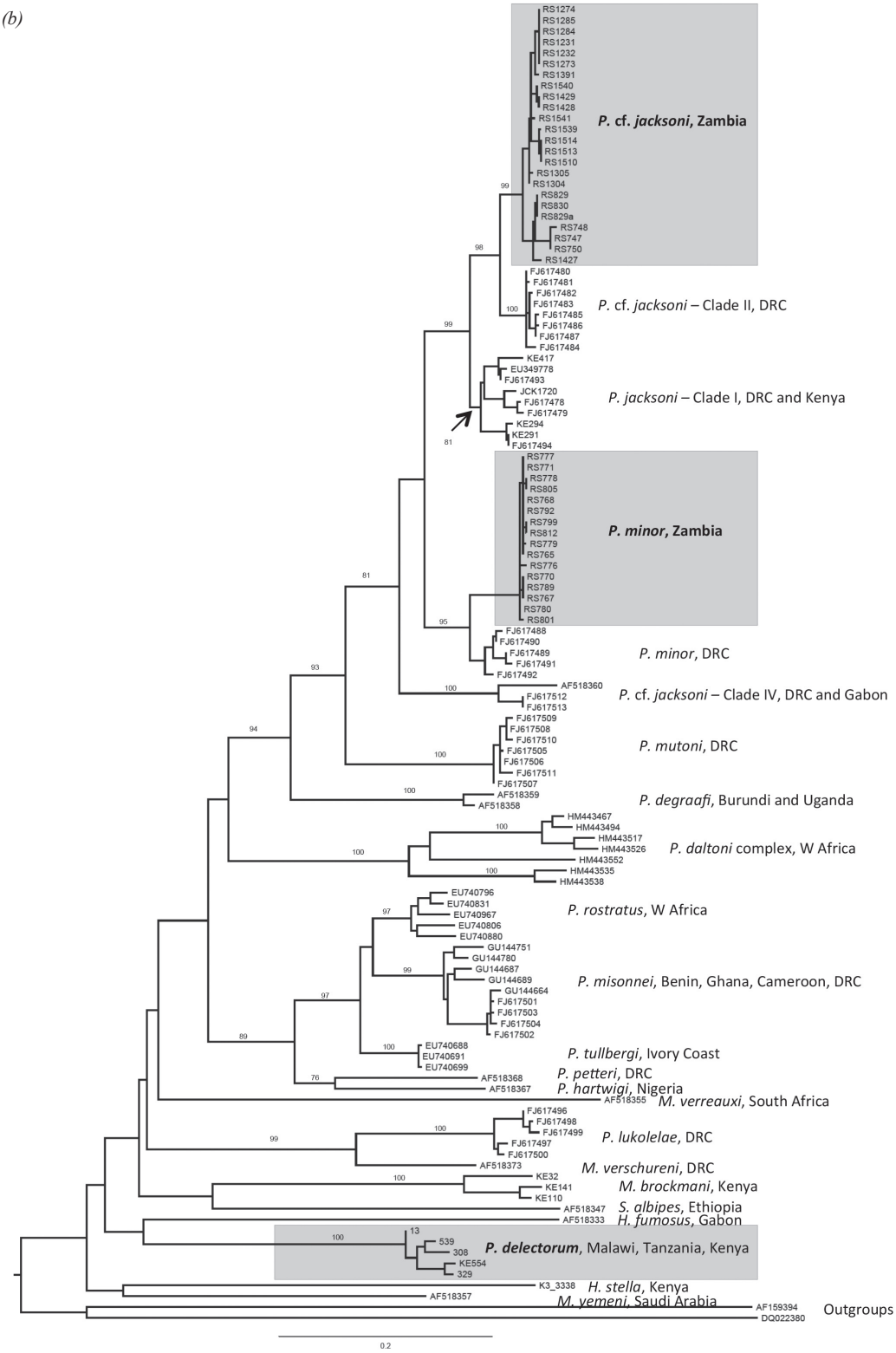


Fig. 3. *Maximum likelihood phylogenetic trees of available sequences of cytochrome* b*. For more details of specimens used see Supplementary Table S1. Species occuring in Zambia are highlighted by grey background. The numbers above the branches represent the bootstrap values (1000 bootstraps in PhyML); only values higher than 70 are shown. (a) genus* Mastomys sensu lato*; (b) genus* Praomys *sensu lato; (c) genus* Hylomyscus*.*

(b)


RS1274
RS1285
RS1284
RS1231
RS1232
RS1273
RS1391
RS1540
RS1429
RS1428
RS1541
RS1539
RS1514
RS1513
RS1510
RS1305
RS1304
RS829
RS830
RS829a
RS748
RS747
RS750
RS1427
P. cf. jacksoni, Zambia
FJ617480
FJ617481
FJ617482
FJ617483
FJ617485
FJ617486
FJ617487
FJ617484
P. cf. jacksoni – Clade II, DRC
KE417
EU349778
FJ617493
JCK1720
FJ617478
FJ617479
KE294
KE291
FJ617494
P. jacksoni – Clade I, DRC and Kenya
RS777
RS771
RS778
RS805
RS768
RS792
RS799
RS812
RS779
RS765
RS776
RS770
RS789
RS767
RS780
RS801
P. minor, Zambia
FJ617488
FJ617490
FJ617489
FJ617491
FJ617492
P. minor, DRC
AF518360
FJ617512
FJ617513
P. cf. jacksoni – Clade IV, DRC and Gabon
FJ617509
FJ617508
FJ617510
FJ617505
FJ617506
FJ617511
FJ617507
P. mutoni, DRC
AF518359
AF518358
P. degraafi, Burundi and Uganda
HM443467
HM443494
HM443517
HM443526
HM443552
P. daltoni complex, W Africa
HM443535
HM443538
EU740796
EU740831
EU740967
EU740806
EU740880
P. rostratus, W Africa
GU144751
GU144780
GU144687
GU144689
GU144664
FJ617501
FJ617503
FJ617504
FJ617502
P. misonnei, Benin, Ghana, Cameroon, DRC
EU740688
EU740691
EU740699
P. tullbergi, Ivory Coast
AF518368
AF518367
P. petteri, DRC
P. hartwigi, Nigeria
AF518355
M. verreauxi, South Africa
FJ617496
FJ617498
FJ617499
FJ617497
FJ617500
P. lukolelae, DRC
AF518373
M. verschureni, DRC
KE32
KE141
KE110
M. brockmani, Kenya
AF518347
S. albipes, Ethiopia
AF518333
H. fumosus, Gabon
13
539
308
KE554
329
P. delectorum, Malawi, Tanzania, Kenya
K3_3338
H. stella, Kenya
AF518357
M. yemeni, Saudi Arabia
AF159394
DQ022380
Outgroups
0.2

individuals from Nyika Plateau, the Makutu Mountains and the Mafinga Mountains from Livingstone Museum (individuals collected between 1969 and 1971; Supplementary Table S1). We failed to obtain complete sequences of cytochrome *b* from museum material, but partial sequences significantly clustered with *P. delectorum* from northern Malawi (not shown). Our phylogenetic analysis (in accordance with other recent analyses, e.g. Kennis et al. 2011, Nicolas et al. 2012) questions the inclusion of this species into the genus *Praomys* (Fig. 3b), but better taxon sampling and sequences from more genes (including nuclear ones) are required for more definitive conclusion.

(c) Other genera

Hylomyscus anselli (Bishop, 1979)

The genus *Hylomyscus* is widely distributed in African tropical forests with the highest diversity in central Africa (Nicolas et al. 2006, Kennis et al., submitted). Central and Western African species were recently reviewed by using sequences of mtDNA and morphometry, two new species were described and other candidates for new species were identified (Nicolas et al. 2006, 2008a, 2010, Kennis et al., submitted). Geographically restricted populations of *Hylomyscus* occur also outside continuous tropical forest areas, e.g. in Zambia and afromontane forests in eastern Africa (e.g. Tanzania, Kenya, Albertine rift). These populations were recently revised by Carleton & Stanley (2005) and Carleton et al. (2006), however only morphological traits were evaluated. The only genetic data from eastern Africa were provided by Huhndorf et al. (2007), which used short fragments of cytochrome *b* gene for analysis of phylogeographical structure of *H.* cf. *denniae* in mountains of the Albertine rift.

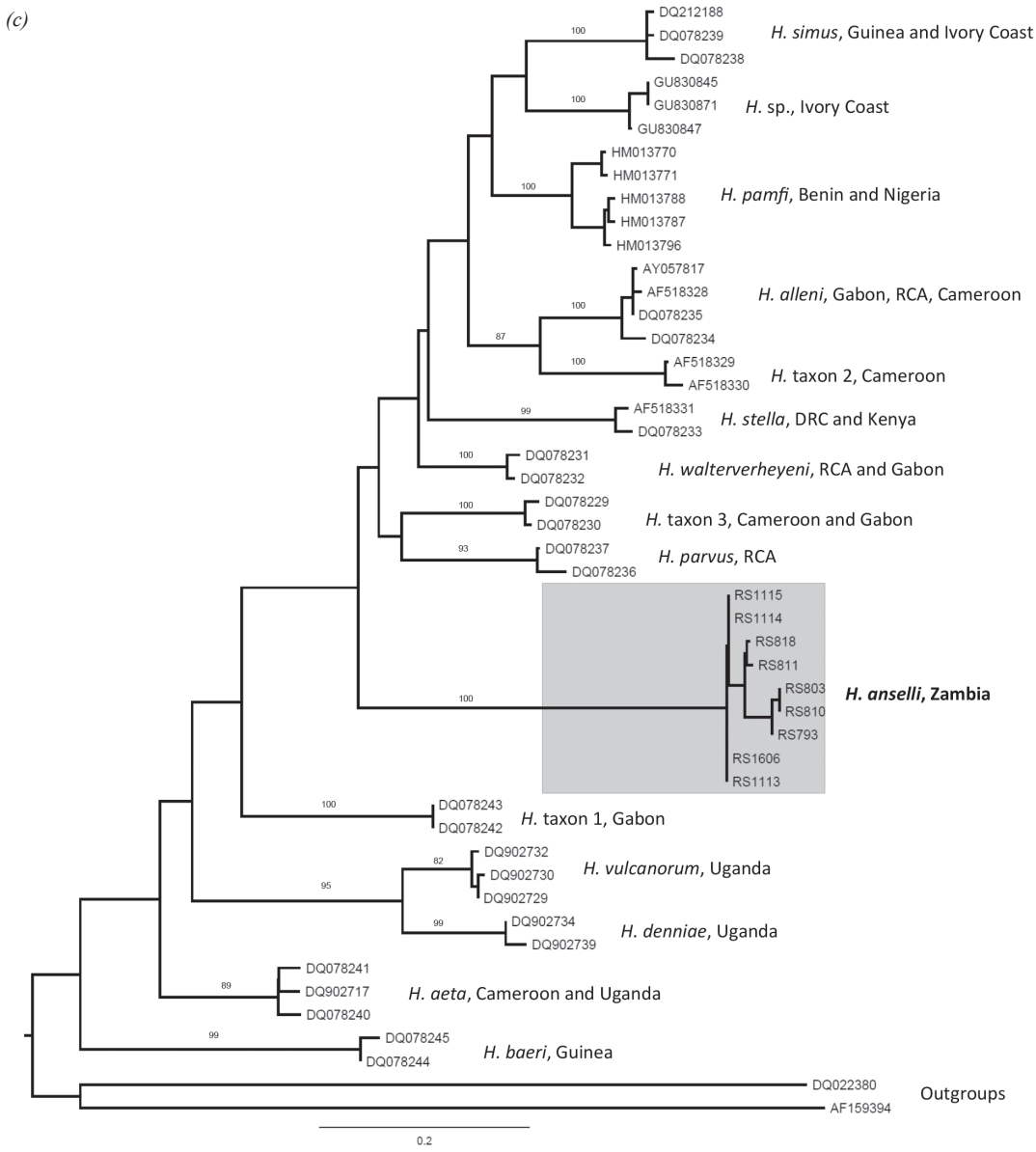
In Zambia, Ansell (1978) reported only one species of the genus (under the name *Praomys* (*Hylomyscus*) *denniae*). Carleton & Stanley (2005) analysed morphological data of *Hylomyscus* from Tanzania and found that a population from Mbizi Mountains in south-western Tanzania is very different from other Tanzanian populations, but very similar to the type series of the subspecies *Praomys* (*Hylomyscus*) *denniae anselli* Bishop, 1979 described from Mwinilunga District in Zambia. Furthermore, both Tanzanian groups are different from *H. denniae* that was described from Uganda. As a consequence, Carleton & Stanley (2005) described a new species *H. arcimontensis* distributed in Mount Rungwe and the Eastern Arc Mountains of central and eastern Tanzania, and *H. anselli* from south-western Tanzania and northern Zambia was elevated to species rank. In our data, *Hylomyscus* has very similar distribution as *P.* cf. *jacksoni*, i.e. in the northern part of the country with a mean annual rainfall higher than 1100 mm, but it seems to be less frequent (Fig. 2d). All sequenced individuals from Zambia cluster in a single group, which is very distinct from all available sequences from previous studies (Fig. 3c). Because no sequences of animals from *H. anselli* species group (i.e. *H. anselli* and *H. arcimontensis*) have been previously published, this suggests that all Zambian individuals belong to *H. anselli*.

Zelotomys hildegardeae (Thomas, 1902)

This species is probably widely, although very sparsely, distributed in northern and especially southern savanna zones. It has been recorded from central Angola, Zambia, northern Malawi, western Uganda, Rwanda, Burundi, eastern DRC, Sudan and Central African Republic. This species is usually associated with moist grassy savanna and scrub, on the edge of swamps and forests and in tall grasslands; however, on the Nyika Plateau in Malawi it has been recorded from pine plantations and near houses (Coetzee & Van der Straeten 2008). In Zambia, this species was found in a relatively large area (Fig. 2d), but always in low numbers, which makes a more precise estimation of its distribution difficult (Ansell 1978). Scarcity of its occurrence is also indicated by its absence in a large collection by Mathias Kawalika, who carried intensive trapping and analysis of owl pellets in Ndola and surroundings (Kawalika 2004), despite Ndola being the type locality for *Z. h. shortridgei* (Ansell 1978). We did not capture any individuals of this species during our recent survey.

Colomys goslingi Thomas & Wroughton, 1907

Colomys goslingi possibly represents a complex of several similar species and further studies are needed to clarify the taxonomic status of populations currently allocated to this species (Kerbis Peterhans et al. 2008). It has been patchily recorded from large areas covered by more or less continuous tropical forest from Liberia in the west, to north-western Zambia in the south and western Ethiopia in the east. The range limits are not fully resolved, but it has not been recorded from the rainforests of Central African Republic or the forests of West Africa (except Gabon and Liberia). The species has been recorded up to 3200 m a.s.l., while the lowest elevation recorded is about 400 m, but generally it occurs above 800 m a.s.l. (Kerbis Peterhans et al. 2008). The only record in Zambia originates from a single lower mandible taken



from an owl pellet in 1964 at Mundwiji Plain (Ansell 1965) (Fig. 2d). Based on the facts that a follow-up expedition failed to collect it there (Ansell 1978), the absence of newer records and a rapid disappearance of suitable habitats, its current presence in Zambia is questionable.

Conclusions

Revised species list of the Praomyini tribe in Zambia

Based on historical records and our recent data, there are eight species that are currently placed in the tribe Praomyini in Zambia. By using analysis of mtDNA sequences, we report for the first time the presence of two new species in Zambia – *Praomys minor* and *Mastomys coucha*. Both species are so-called cryptic species of two more widely distributed species – *P.* cf. *jacksoni* and *M. natalensis*. Both species pairs can also be distinguished by detailed morphological analysis; however such analysis is still missing for Zambian populations. Here these two new species for Zambia reach the border of their distribution and probably occur only in a very limited number of localities.

The taxonomic status of three previously reported species was clarified based on genetic data. It was confirmed that *M. natalensis* is widely distributed all over the country, while two additional species are restricted to the northern part of the country with higher precipitation and remnants of forest habitats. First, we report the first sequences of *H. anselli*, which are very distinct from all other available *Hylomyscus*. Second, *P.* cf. *jacksoni* occurring in Zambia is genetically distinct from nominotypical *P. jacksoni* described from Uganda and probably represent a new species worth of further taxonomical studies (Kennis et al. 2011).

The remaining three species were not captured during our recent field work. *P. delectorum* was, however, confirmed genetically by analysis of museum samples. This species is in Zambia geographically limited to Nyika Plateau and neighbouring mountain ranges, and we captured it frequently in nearby localities in Malawi. *Z. hildegardae* was not found during our fieldwork, probably because of its relative rarity. It is reported to occur at several localities in Zambia but always in low densities or (alternatively) it is very trap-shy. The presence of the last Praomyini species, *C. goslingi*, in Zambia is questionable. The only record originates from owl pellets and is based on a single individual (Ansell 1965). The species has never been captured in Zambia and because it is a typical species of tropical forest, i.e. the habitat largely destroyed by humans in recent decades, we think that the probability that it is still a part of the mammal fauna of Zambia is very low.

Even if the research of small mammals in Zambia has been relatively intense in the past (especially in the middle of the last century), the list of Praomyini is not necessarily complete. Genetic data led to description of many new species in this group throughout Africa and it is possible that some of them could also occur in Zambia. (1) It concerns especially the species-rich genera *Praomys* and *Hylomyscus*, which are particularly species-rich in the tropical forests of the Congo Basin (e.g. Nicolas et al. 2005, Kennis et al. 2011) and the distribution of some of them may reach northern Zambia. (2) Endemic species of the Tanzanian mountains, *H. arcimontensis*, occurs in the Rungwe Mountains, which shares many species with Nyika Plateau (e.g. *P. delectorum*) and any *Hylomyscus* potentially found at this area will be very interesting. (3) The northernmost part of Zambia is also worthy of future faunistic studies, because it is relatively close to the Albertine rift, which is known for its very high rodent endemicism. For example, two endemic species of *Hylomyscus* (*H. denniae*, *H. vulcanorum*) and one of *Praomys* (*P. degraafi*) are restricted only to mountain forests of the Albertine rift and they could have been neglected in previous studies of Zambian fauna, because they belong to complexes of cryptic species. (4) Another candidate for enriching Zambian fauna is *M. shortridgei*. Its presence in north-western Botswana (Ansell 1978, Coetzee & Griffin 2008) suggests its possible occurrence also in Zambia. Our effort to capture this cryptic species in Zambia was not successful; however, more intensive field research is necessary in the south-western part of the country.

Distribution of species from the Praomyini tribe in Zambia – biogeographical and conservation implications

All but one of the studied species of Praomyini are not widely distributed in the whole country. The only exception is *M. natalensis* and its occurrence in the whole country is probably at least partly related to its opportunistic habitat preferences and partial synanthropy. The remaining species, however, can be used as more general markers of important biogeographic areas in Zambia.

(a) Four Praomyini species are distributed in the northern part of the country, where mean annual rainfall is higher than cca 1100 mm (Fig. 2b, c). Two of them (*P.* cf. *jacksoni*, *H. anselli*) are relatively frequent in many forest habitats, but the remaining two (*P. minor*, *C. goslingi*) were recorded only in isolated forest habitats in the Mwinilunga District in north-western Zambia. This region would surely deserve to be given more attention and support by nature conservation efforts. Kawalika (2004) reviewed data from Ansell (1978) and found the highest diversity (31-33 species) of rodents in Zambia in this area. Based on Ansell (1978), there are also other rodents of humid tropical forest, whose distribution in Zambia is restricted to the Mwinilunga region (e.g. *Pelomys minor*, *Hybomys univittatus* and *Malacomys longipes*). Similar patterns also exist for other animals; for example one of Zambian endemic mammals, Ansell's musk shrew (*Crocidura ansellorum*), occurs only here (Hutterer & Dippenaar 1987). This relatively small area also hosts unique larger mammal fauna, which are found hardly anywhere else in Zambia. Among the most conspicuous, there are records of lowland forest mammals such as Angola pied-colobus (*Colobus angolensis*), red tailed monkey (*Cercopithecus ascanius*), African palm civet (*Nandinia binottata*), tree pangolin (*Manis tricuspis*), Thomas's bushbaby (*Galagoides thomasi*), giant otter shrew (*Potamogale velox*), and hammer bat (*Hypsignathus monstrosus*)

(Ansell 1978, Cotterill 2002, Leonard 2005, Van Daele in litt.). Very close faunal affinity with Guinean-Congolian biome is best documented in birds, because almost all Guinean-Congolian elements occurring in Zambia live just here (Leonard 2005). Unfortunately, this unique area, which should be among the most important conservation priorities in Zambia, is not included in any protected areas system. The only protection is guaranteed by owners of private commercial farms (Nchila Wildlife Reserve) or National Heritage Conservation Commission (Source of Zambezi National Monument). Adjacent areas without any protection are threatened especially by forest clearance (Cotterill 2002, Leonard 2005). According to Cotterill (2002), species richness and diversity here is higher than elsewhere in Zambia or southern Africa south of 10° S, which reinforces arguments for its conservation as biodiversity hotspot of regional and global importance.

(b) The Nyika Plateau and neighbouring mountains (Makutus and Mafingas) make up an afromontane region, which was identified as a rodent "biodiversity hotspot" (Kawalika 2004). This is consistent with the expectation based on the knowledge of high biodiversity in the Malawian part of the Nyika Plateau, just across the border (cf. Chitaukali et al. 2001 and literature cited therein), although sampling on the Zambian site was rather incidental and punctual both in space and time (Ansell 1978). This is the only truly montane region of Zambia belonging to the so-called "Southern Rift Mountains", i.e. the biogeographical area with many endemics (e.g. Taylor et al. 2009, Bayliss et al. 2010). There are several other rodent species, recorded in Zambia only from this area, e.g. undescribed species from *Otomys lacustris* complex (in Ansell 1978 as *O. denti*, in Taylor et al. 2009 as *O. sungae*), *Otomys uzungwensis* (in Ansell 1978 as *O. typus*), *Paraxerus lucifer*, *Beamys hindei*, *Rhabdomys dilectus*) (Ansell 1978 and subsequent actualizations in Rambau et al. 2003, Taylor et al. 2009, 2011). Most of the large mammals were heavily poached out in those mountain areas. Small mammals are threatened mainly by habitat destruction. For example, the area of a mature montane forest, i.e. the habitat preferred by *Praomys delectorum*, is estimated to be only about 200 ha in total making vertebrate species restricted to this specific habitat vulnerable on a national scale (Leonard 2005).

(c) The south-western part of the country (mainly to the west and southwest of the River Zambezi) is characterised by lower annual rainfall (less than 700 mm). It lies mainly on Kalahari sands and ideally belongs to the Southern African subregion (usually limited in the north by the River Zambezi). Besides Southern African large mammals, also several rodents occur in Zambia exclusively in this area, e.g. *M. coucha*, *Fukomys damarensis*, *Gerbilliscus* cf. *brantsii* and *Mus setzeri* have northern limits of their distribution here (Ansell 1978, this study and our unpublished data). Major habitats here are forest, grasslands and savannah woodland dominated with *Acacia* trees. The main conservation problem here is a rapid decrease and degradation of forest dominated by Zambezi teak (*Baikiaea plurijuga*) by illegal timber extraction and burning (Leonard 2005). The small mammal community is poorly known in this area and further sampling could reveal the presence of other species with confirmed distribution in neighbouring countries, such as *Mastomys shortridgei*, *Mus indutus* or *Gerbillurus paeba*.

Acknowledgements

The work was supported by the Czech Science Foundation (project no. P506/10/0983), Ministry of Education, Youth and Sports (project no. 6007665801), and Grant Agency of University of South Bohemia (project no. 135/2010/P). We are grateful to the Zambia Wildlife Authority (ZAWA) for permission to carry out our field work, especially to Wilbroad Chansa and Chuma Simukonda from ZAWA Chilanga Head Office for assistance in organising of permits. We are also obliged to the management of Kasanka Trust Ltd., African Parks, Mutinondo Wildneress, and many private owners for allowing us to sample small mammals in their fields, especially Bruce-Miller and Fisher families (in Bruce-Miller Game Farm and Nchila Wildlife Reserve). Our thank belongs to all local ZAWA officers and ZAWA rangers in national parks and other protected areas, together with many local authorities, who assisted us with our research. V. Nicolas and M. Macholán provided useful comments to previous version of the manuscript. Last but not least, we are grateful to Ema Hrouzková, Ondřej Kott, Matěj Lövy, and Jan Ritter, who assisted us with trapping of small mammals in the field.

Literature

Akpatou B., Nicolas V., Pires D., N'Goran E. & Colyn M. 2007: Morphometric differentiation between two murid rodents, *Praomys tullbergi* (Thomas, 1894) and *Praomys rostratus* (Miller, 1900), in West Africa. *Zootaxa 1607: 21–34.*

Ansell W.F.H. 1965: Addenda and corrigenda to "Mammals of Northern Rhodesia" No. 2. *Puku 3: 1–14.*
Ansell W.F.H. 1978: The mammals of Zambia. *Zambia National Parks & Wildlife Service, Chilanga.*
Bayliss J., Burrow C., Martell S. & Staude H. 2010: An ecological study of the relationship between two living fossils in Malawi: the Mulanje tiger moth (*Callioratis grandis*) and the Mulanje cycad (*Encephalartos gratus*). *Afr. J. Ecol. 48: 472–480.*
Briese T., Paweska J.T., McMullan L.K., Hutchison S.K., Street C., Palacios G., Khristova M.L., Weyer J., Swanepoel R., Egholm M., Nichol S.T. & Lipkin W.I. 2009: Genetic detection and characterization of Lujo virus, a new hemorrhagic fever – associated arenavirus from Southern Africa. *PLoS Pathog. 5: e1000455.*
Bryja J., Granjon L., Dobigny G., Patzenhauerová H., Konečný A., Duplantier J.M., Gauthier P., Colyn M., Durnez L., Lalis A. & Nicolas V. 2010: Plio-Pleistocene history of West African Sudanian savanna and the phylogeography of the *Praomys daltoni* complex (Rodentia): the environment/geography/genetic interplay. *Mol. Ecol. 19: 4783–4799.*
Burda H., Zima J., Scharff A., Macholán M. & Kawalika M. 1999: The karyotypes of *Cryptomys anselli* sp. nova and *Cryptomys kafuensis* sp. nova: new species of the common mole-rat from Zambia (Rodentia, Bathyergidae). *Z. Säugetierkd. 64: 36–50.*
Carleton M.D., Kerbis Peterhans J.C. & Stanley W. 2006: Review of the *Hylomyscus denniae* group (Rodentia: Muridae) in Eastern Africa, with comments on the generic allocation of *Epimys endorobae* Heller. *P. Biol. Soc. Wash. 119: 293–325.*
Carleton M.D. & Stanley W.T. 2005: Review of the *Hylomyscus denniae* complex (Rodentia: Muridae) in Tanzania, with a description of a new species. *P. Biol. Soc. Wash. 118: 619–646.*
Carleton M.D. & Stanley W.T. 2012: Species limits within the *Praomys delectorum* group (Rodentia: Muridae: Murinae) of East Africa: a morphometric reassessment and biogeographical implications. *Zool. J. Linn. Soc. 165: 420–469.*
Castiglia R., Corti M., Colangelo P., Annesi F., Capanna E., Verheyen W., Sichilima A. & Makundi R. 2003: Chromosomal and molecular characterization of *Aethomys kaiseri* from Zambia and *Aethomys chrysophilus* from Tanzania (Rodentia, Muridae). *Hereditas 139: 81–89.*
Castiglia R., Gornung E. & Corti M. 2002: Cytogenetic analyses of chromosomal rearrangements in *Mus minutoides/musculoides* from North-West Zambia through mapping of the telomeric sequence ($(TTAGGG)_n$ and banding techniques. *Chromosome Res. 10: 399–406.*
Chidumayo E.N. 1977: The ecology of the single striped grass mouse, *Lemniscomys griselda*, in Zambia. *Mammalia 41: 411–418.*
Chidumayo E.N. 1979: Ecological relations among small rodents inhabiting grasslands at Livingstone, Zambia. *Zambia J. Sci. Technol. 4: 12–20.*
Chidumayo E.N. 1980a: Ecology of rodents at an old quarry in Zambia. *S. Afr. J. Zool. 15: 44–49.*
Chidumayo E.N. 1980b: Population ecology of *Tatera leucogaster* (Rodentia) in southern Zambia. *J. Zool. 190: 325–335.*
Chidumayo E.N. 1984: Observations on populations of multimammate mice at Livingstone, Zambia. *Mammalia 48: 363–376.*
Chitaukali W.N., Burda H. & Kock D. 2001: On small mammals of the Nyika Plateau, Malawi. *African Small Mammals*: *415–426.*
Coetzee N. & Griffin M. 2008: *Mastomys shortridgei*. In: IUCN 2011. IUCN Red List of Threatened Species. *Version 2011.2.*
Coetzee N., Griffin M. & Granjon L. 2008: *Mastomys coucha*. In: IUCN 2011. IUCN Red List of Threatened Species. *Version 2011.2.*
Coetzee N. & Van der Straeten E. 2008: *Zelotomys hildegardeae*. In: IUCN 2011. IUCN Red List of Threatened Species. *Version 2011.2.*
Colangelo P., Leirs H., Castiglia R., D'aes M., Meheretu Y. & Verheyen E. 2010: New data on the distribution and phylogenetic position of *Mastomys awashensis* (Rodentia, Muridae). *Mamm. Biol. 75: 459–462.*
Corti M., Castiglia R., Annesi F. & Verheyen W. 2004: Mitochondrial sequences and karyotypes reveal hidden diversity in African pouched mice (subfamily Cricetomyinae, genus *Saccostomus*). *J. Zool. 262: 413–424.*
Corti M., Castiglia R., Colangelo P., Capanna E., Beolchini F., Bekele A., Oguge N., Makundi R., Sichilima A. & Leirs H. 2005: Cytotaxonomy of rodent species from Ethiopia, Kenya, Tanzania and Zambia. *Belg. J. Zool. 135: 197–216.*

Cotteril F.P.D. 2002: Notes on mammal collections and biodiversity conservation in the Ikelenge pedicle, Mwinilunga district, Northwestern Zambia. *Biodiversity Foundation for Africa, Bullawayo. Occasional Publication in Biodiversity 10: 1–18.*

Coulibaly-N'Golo D., Allali B., Kouassi S.K., Fichet-Calvet E., Becker-Ziaja B., Rieger T., Ölschläger S., Dosso H., Denys C., ter Meulen J., Akoua-Koffi C. & Günther S. 2011: Novel arenavirus sequences in *Hylomyscus sp.* and *Mus (Nannomys) setulosus* from Côte d'Ivoire: implications for evolution of arenaviruses in Africa. *PloS ONE 6: e20893.*

Dobigny G., Lecompte E., Tatard C., Gauthier P., Ba K., Denys C., Duplantier J. & Granjon L. 2008: An update on the taxonomy and geographic distribution of the cryptic species *Mastomys kollmannspergeri* (Muridae, Murinae) using combined cytogenetic and molecular data. *J. Zool. 276: 368–374.*

Dobigny G., Tatard C., Kane M., Gauthier P., Brouat C., Ba K. & Duplantier J.M. 2011: A cytotaxonomic and DNA-based survey of rodents from Northern Cameroon and Western Chad. *Mamm. Biol. 76: 417–427.*

Fichet-Calvet E., Lecompte E., Koivogui L., Daffis S. & ter Meulen J. 2008: Reproductive characteristics of *Mastomys natalensis* and Lassa virus prevalence in Guinea, West Africa. *Vector-Borne Zoonotic Dis. 8: 41–48.*

Fichet-Calvet E., Lecompte E., Koivogui L., Soropogui B., Doré A., Kourouma F., Sylla O., Daffis S., Koulémou K. & ter Meulen J. 2007: Fluctuation of abundance and Lassa virus prevalence in *Mastomys natalensis* in Guinea, West Africa. *Vector-Borne Zoonotic Dis. 7: 119–128.*

Granjon L., Cosson J., Quesseveur E. & Sicard B. 2005: Population dynamics of the multimammate rat *Mastomys huberti* in an annually flooded agricultural region of central Mali. *J. Mammal. 86: 997–1008.*

Granjon L., Duplantier J., Catalan J. & Britton-Davidian J. 1997: Systematics of the genus *Mastomys* (Thomas, 1915) (Rodentia: Muridae). A review. *Belg. J. Zool. 127: 7–18.*

Gratz N.G. 2006: Vector- and rodent-borne diseases in Europe and North America: distribution, public health burden, and control. *Cambridge University Press, Cambridge.*

Guindon S., Lethiec F., Duroux P. & Gascuel O. 2005: PHYML online – a web server for fast maximum likelihood-based phylogenetic inference. *Nucleic Acids Res. 33: W557–W559.*

Hatt R.T. 1934: Fourteen hitherto unrecognized African rodents. *American Museum Novitates 708: 1–15.*

Hoffmann M., Grubb P., Groves C.P., Hutterer R., Van der Straeten E., Simmons N. & Bergmans W. 2009: A synthesis of African and western Indian Ocean Island mammal taxa (Class: Mammalia) described between 1988 and 2008: an update to Allen (1939) and Ansell (1989). *Zootaxa 2205: 1–36.*

Huhndorf M.H., Kerbis Peterhans J.C. & Loew S.S. 2007: Comparative phylogeography of three endemic rodents from the Albertine Rift, east central Africa. *Mol. Ecol. 16: 663–674.*

Hutterer R. & Dippenaar N.J. 1987: A new species of *Crocidura* Wagler, 1832 (Soricidae) from Zambia. *Bonn. Zool. Beitr. 38 (1): 1–7.*

Ishii A., Thomas Y., Moonga L., Nakamura I., Ohnuma A., Hang'ombe B., Takada A., Mweene A. & Sawa H. 2011: Novel arenavirus, Zambia. *Emerg. Infect. Dis. 17: 1921.*

Jachmann H. 2000: Zambia's wildlife resources: a brief ecology. *Wildlife Resource Monitoring Unit, Environmental Council of Zambia, Lusaka, Zambia.*

Kan Kouassi S., Nicolas V., Aniskine V., Lalis A., Cruaud C., Couloux A., Colyn M., Dosso M., Koivogui L. & Verheyen E. 2008: Taxonomy and biogeography of the African pygmy mice, Subgenus *Nannomys* (Rodentia, Murinae, *Mus*) in Ivory Coast and Guinea (West Africa). *Mammalia. 72: 237–252.*

Kawalika M. 2004: Rodents of Ndola (Copperbelt Province, Zambia). *Ph.D. Thesis. University of Duisburg-Essen.*

Kennis J., Nicolas V., Hulselmans J., Katuala P.G.B., Wendelen W., Verheyen E., Dudu A.M. & Leirs H. 2011: The impact of the Congo River and its tributaries on the rodent genus *Praomys*: speciation origin or range expansion limit? *Zool. J. Linn. Soc. 163: 983–1002.*

Kerbis Peterhans J.C., Schlitter D., Dieterlen F. & Cotterill F. 2008: *Colomys goslingi*. In: IUCN 2011. IUCN Red List of Threatened Species. *Version 2011.2.*

Lecompte E., Aplin K., Denys C., Catzeflis F., Chades M. & Chevret P. 2008: Phylogeny and biogeography of African Murinae based on mitochondrial and nuclear gene sequences, with a new tribal classification of the subfamily. *BMC Evol. Biol. 8, 199.*

Lecompte E., Brouat C., Duplantier J., Galan M., Granjon L., Loiseau A., Mouline K. & Cosson J. 2005: Molecular identification of four cryptic species of *Mastomys* (Rodentia, Murinae). *Biochem. Syst. Ecol. 33: 681–689.*

Lecompte E., Fichet-Calvet E., Daffis S., Koulemou K., Sylla O., Kourouma F., Dore A., Soropogui B., Aniskin V., Allali B., Kan S., Lalis A., Koivogui L., Gunther S., Denys C. & ter Meulen J. 2006: *Mastomys natalensis* and Lassa fever, West Africa. *Emerg. Infect. Dis. 12: 1971–1974.*

Lecompte E., Granjon L., Kerbis Peterhans J.C. & Denys C. 2002: Cytochrome *b*-based phylogeny of the *Praomys* group (Rodentia, Murinae): a new African radiation? *C. R. Biol. 325: 827–840.*

Leonard P.M. 2005: Important bird areas in Zambia – priority sites for conservation. *The Zambian Ornithological Society, Lusaka.*

Mayaux P., Bartholome' E., Fritz S. & Belward A. 2004: A new land-cover map of Africa for the year 2000. *J. Biogeogr. 31: 861–877.*

Mullin S., Pillay N. & Taylor P. 2004: Cranial variation and geographic patterns within the *Dasymys rufulus* complex (Rodentia: Muridae). *J. Mammal. 85: 911–923.*

Mullin S., Pillay N. & Taylor P. 2005: The distribution of the water rat *Dasymys* (Muridae) in Africa: a review. *S. Afr. J. Sci. 101: 117–124.*

Musser G. & Carleton M. 2005: Superfamily Muroidea. In: Wilson D.E. & Reeder D.M. (eds.), Mammal species of the World, Vol. 2. *John Hopkins University Press, Baltimore: 894–1531.*

Nicolas V., Bryja J., Akpatou B., Konečný A., Lecompte E., Colyn M., Lalis A., Couloux A., Denys C. & Granjon L. 2008b: Comparative phylogeography of two sibling species of forest-dwelling rodent (*Praomys rostratus* and *P. tullbergi*) in West Africa: different reactions to past forest fragmentation. *Mol. Ecol. 17: 5118–5134.*

Nicolas V., Missoup A.D., Denys C., Kerbis Peterhans J.C., Katuala P., Couloux A. & Colyn M. 2011: The roles of rivers and Pleistocene refugia in shaping genetic diversity in *Praomys misonnei* in tropical Africa. *J. Biogeogr. 38: 191–207.*

Nicolas V., Olayemi A., Wendelen W. & Colyn M. 2010: Mitochondrial DNA and morphometrical identification of a new species of *Hylomyscus* (Rodentia: Muridae) from West Africa. *Zootaxa 2579: 30–44.*

Nicolas V., Quérouil S., Verheyen E., Verheyen W., Mboumba J., Dillen M. & Colyn M. 2006: Mitochondrial phylogeny of African wood mice, genus *Hylomyscus* (Rodentia, Muridae): implications for their taxonomy and biogeography. *Mol. Phylogenet. Evol. 38: 779–793.*

Nicolas V., Schaeffer B., Missoup A.D., Kennis J., Colyn M., Denys C., Tatard C., Cruaud C. & Laredo C. 2012: Assessment of three mitochondrial genes (16S, Cytb, CO1) for identifying species in the Praomyini tribe (Rodentia: Muridae). *PLoS ONE 7: e36586.*

Nicolas V., Verheyen E., Verheyen W., Hulselmans J., Dillen M., Akpatou B., Dudu A., Wendelen W. & Colyn M. 2005: Systematics of African lowland rainforest *Praomys* (Rodentia, Muridae) based on molecular and craniometrical data. *Zool. J. Linn. Soc. 145: 539–553.*

Nicolas V., Wendelen W., Barriere P., Dudu A. & Colyn M. 2008a: Morphometric variation in *Hylomyscus alleni* and *H. stella* (Rodentia: Muridae), and description of a new species. *J. Mammal. 89: 222–231.*

Olson D.M., Dinerstein E., Wikramanayake E.D., Burgess N.D., Powell G.V.N., Underwood E.C., D'amico J.A., Itoua I., Strand H.E., Morrison J.C., Loucks C.J., Allnutt T.F., Ricketts T.H., Kura Y., Lamoreux J.F., Wettengel W.W., Hedao P. & Kassem K.R. 2001: Terrestrial ecoregions of the world: a new map of life on earth. *BioScience 51: 933–938.*

Orba Y., Kobayashi S., Nakamura I., Ishii A., Hang'ombe B.M., Mweene A.S., Thomas Y., Kimura T. & Sawa H. 2011: Detection and characterization of a novel polyomavirus in wild rodents. *J. Gen. Virol. 92: 789.*

R Development Core Team 2011: R: a language and environment for statistical computing. *R Foundation for Statistical Computing, Vienna, Austria. http://www.R-project.org/*

Rambau R., Robinson T. & Stanyon R. 2003: Molecular genetics of *Rhabdomys pumilio* subspecies boundaries: mtDNA phylogeography and karyotypic analysis by fluorescence in situ hybridization. *Mol. Phylogenet. Evol. 28: 564–575.*

Singleton G.R., Belmain S.R., Brown P.R. & Hardy B. 2010: Rodent outbreaks: ecology and impacts. *International Rice Research Institute, Los Baños (Philippines).*

Sluydts V., Davis S., Mercelis S. & Leirs H. 2009: Comparison of multimammate mouse (*Mastomys natalensis*) demography in monoculture and mosaic agricultural habitat: implications for pest management. *Crop Prot. 28: 647–654.*

Taylor P.J., Maree S., Sandwyk J.V., Kerbis Peterhans J.C., Stanley W.T., Verheyen E., Kaliba P., Verheyen W., Kaleme P. & Bennett N.C. 2009: Speciation mirrors geomorphology and palaeoclimatic history in African

laminate-toothed rats (Muridae: Otomyini) of the *Otomys denti* and *Otomys lacustris* species-complexes in the "Montane Circle" of East Africa. *Biol. J. Linn. Soc. 96: 913–941.*

Taylor P.J., Lavrenchenko L.A., Carleton M.D., Verheyen E., Bennett N.C., Oosthuizen C.J. & Maree S. 2011: Specific limits and emerging diversity patterns in East African populations of laminate-toothed rats, genus *Otomys* (Muridae: Murinae: Otomyini): revision of the *Otomys typus* complex. *Zootaxa 3024: 1–66.*

Van Daele P.A.A.G., Verheyen E., Brunain M. & Adriaens D. 2007: Cytochrome *b* sequence analysis reveals differential molecular evolution in African mole-rats of the chromosomally hyperdiverse genus *Fukomys* (Bathyergidae, Rodentia) from the Zambezian region. *Mol. Phylogenet. Evol. 45: 142–157.*

Van der Straeten E. 2008: Notes on the *Praomys* of Angola with the description of a new species (Mammalia: Rodentia: Muridae). *Stuttgart. Beitr. Naturk. (A) 1: 123–131.*

Verheyen W., Hulselmans J., Dierckx T., Colyn M., Leirs H. & Verheyen E. 2003: A craniometric and genetic approach to the systematics of the genus *Dasymys* Peters, 1875, selection of a neotype and description of three new taxa (Rodentia, Muridae, Africa). *Bull. Inst. R. Sci. Nat. Belg. Biol. 73: 27–71.*

Verheyen W., Hulselmans J., Wendelen W., Leirs H., Corti M., Backeljau T. & Verheyen E. 2011: Contribution to the systematics and zoogeography of the East-African *Acomys spinosissimus* Peters, 1852 species complex and the description of two new species (Rodentia: Muridae). *Zootaxa 3059: 1–35.*

White F. 1983: The Zambezian regional centre of endemisms. In: The vegetation of Africa. *UNESCO, Paris: 86–101.*

Supplementary online materials

Supplementary Table S1. *(a) List of sequenced individuals and their GenBank Accession numbers; (b) list of sequences retrieved from GenBank used in phylogenetic analyses (Excel file; URL: http://www.ivb.cz/folia/download/bryja_et_al._supplementary_table_s1.xlsx).*

Conclusion

and

future perspective

Molecular genetic analyses detected considerably higher diversity of small mammals than previously expected. Several candidates for new species have been suggested and many new localities for known species have been discovered.

Observed phylogeographic structure revealed geographical barriers to gene flow and current contact zones between neighboring taxonomical units (Zambezi-Kafue river system, Rukwa Rift, Eastern Arc Mountains, Limpopo river valley, etc.).

Molecular dating, evidence for expansion and contraction of population size, and species distribution modeling suggest connection of main lineage splits with rapid fluctuations in forest/savanna biomes during the Late Pleistocene

The main findings and implications of the thesis based on four manuscripts are summarized below:

In **Chapter I** we described phylogeographic structure of the savanna-woodland adapted rodent *Aethomys chrysophilus* and we identified geomorphological factors and climatic processes responsible for the current distribution of its mitochondrial lineages. Seven parapatric clades of *A. chrysophilus* were identified in the mitochondrial CYTB phylogeny.

Phylogeographic approach combined with species distribution modeling and morphometric analysis revealed sporadically overlapping contact zones positioned on previously described biogeographical divides (Zambezi-Kafue river system, Rukwa Rift, and the Eastern Arc Mountains). One internal clade corresponded to populations previously described as a distinct species, *Aethomys ineptus*.

The whole *A. chrysophilus* complex originating approximately 1.3 Mya, with *A. ineptus* split 0.7 Mya. The internal position of *A. ineptus* was confirmed in phylogenetic reconstruction based on

two nuclear genes and thus it is not a consequence of mitochondrial introgression. In addition, skull form variation found no significant difference between *A. ineptus* and the rest of *A. chrysophilus* complex.

In **Chapter II** we described phylogenetic structure and geographical distribution of lineages using mitochondrial, autosomal, and sex chromosome data in another savanna-woodland adapted rodent *Gerbilliscus leucogaster*.

The most recent common ancestor for all *G. leucogaster* lineages occurred during the early Pleistocene. Six mitochondrial lineages that diverged ~0.270-0.100 Mya were documented. Factors most probably responsible for geographical isolation of the lineages were alterations to the course of the Zambezi River and its tributaries and regional 'megadroughts'.

The most widespread lineage exhibited a demographic expansion ~0.065-0.035 Mya, a time that coincides with savanna–woodland expansion across southern Africa. Multilocus autosomal perspective revealed an influence of the Kafue River as a current barrier to gene flow.

In **Chapter III** we performed comprehensive genetic analysis of over 650 individuals *Nannomys*, subgenera of genus *Mus* and important models for the study of sex determination in mammals, collected at approx. 300 localities across the sub-Saharan Africa. Phylogenetic reconstructions based on mitochondrial (CYTB) and nuclear (IRBP) genes identified five species groups and three monotypic ancestral lineages.

We provide evidence for important cryptic diversity and we defined and mapped the distribution of 27 taxonomical units that may correspond to potential new species. Biogeographical

reconstructions based on data spanning all of Africa modified the previous evolutionary scenarios.

First divergences occurred in Eastern African mountains soon after the colonization of the continent and the remnants of these old divergences still occur there, represented by long basal branches of *Mus imberbis* (previously *Muriculus imberbis*) and two undescribed species from Ethiopia and Malawi.

The radiation in drier lowland habitats associated with the decrease of body size is much younger, occurred mainly in a single lineage (minutoides group), and was probably linked to aridification and climatic fluctuations in the middle Pliocene/Pleistocene.

In **Chapter IV** we review available historical records and revise our recently collected material by sequencing the mitochondrial CYTB gene. The presence of eight species of Praomyini in Zambia is documented and the pattern of their geographical distribution is described and discussed.

Three species of predominantly forest dwelling genus *Praomys* were identified, with first record of *P. minor* for Zambia extending its distribution range by more than 1300 km to the south. Another member of the genus *P. cf. jacksoni* was collected in areas with more 1100 mm annual rainfall in the northern part of the country and could represent a new species. Very similar distribution had individuals belonged to genus *Hylomyscus* (possibly *H. anselli*) with all sequenced individuals cluster in a single group distinct from other published sequences.

We have identified important biogeographic areas of Zambia: a) Mwinilunga District in north-western Zambia as area with known faunal affinity with Guinean-Congolian biome; b) the Zambian side of afromontane Nyika Plateau, earlier identified as a rodent "biodiversity hotspot"; c) the south-western part of the country,

mainly southwest of the River Zambezi, characterized < 700 mm annual rainfall.

The results document the versatily of the DNA barcoding for description of species diversity in taxonomically complicated groups with many cryptic species and despite considerable gaps in our data set, we have now enough material to employ comparative phylogeography approaches, required to confirm our observations have broader implications and can be considered pattern for (small) mammals in general.

The species distribution modelling can fill in the gaps in specimen collections and could be used to identify historical distributional patterns, especially possible refugia during unfavorable conditions. Detailed studies on multiple sympatric taxa will provide insight into hybridization patterns on contact zones. And innovations in genetic sequencing technology will provide capacity to reveal the forces shaping complex patterns of genetic variation between individuals, populations, and species. All together these approaches will greatly broaden our understanding of the nature of the factors shaping current, as well as historical, biodiversity patterns in East Africa.

Appendix

Curriculum Vitae

Date of Birth: January 22nd, 1980

Nationality: Czech

Education

Ph.D. studies in Zoology — Since 2008

University of South Bohemia, Faculty of Science

Mgr. in Zoology — 2005 – 2008

University of South Bohemia, Faculty of Science

Bc. in Biology — 2003 – 2005

University of South Bohemia, Faculty of Biological Sciences

Work experience (academic positions):

IVB AVCR, Brno, CZ (Junior Researcher) — 2015–2016

Evolution of rodents in open habitats of East Africa

The Czech Academy of Sciences, Institute of Vertebrate Biology, Květná 8, 603 65 Brno

PŘF JČU, České Budějovice, CZ (Junior Researcher) — 2008 – 2016

Phylogeography of African small mammals

University of South Bohemia; Faculty of Science; Department of Zoology, Branišovská 1760, 370 05

IVB AVCR, Brno, CZ (Junior Researcher) — 2007 – 2008

Nesting behavior of chimpanzees.

The Czech Academy of Sciences, Institute of Vertebrate Biology, Květná 8, 603 65 Brno

PUBLICATIONS

Phylogeography of small African mammals

Krásová J., Mikula O., **Mazoch V.**, Bryja J., Říčan O. & Šumbera R. Evolution of the Grey-bellied pygmy mouse group: highly structured molecular diversity with predictable geographic ranges but morphological crypsis. *Molecular Phylogenetics and Evolution*, 130, 143-155. 2019.

Mazoch V., Mikula O., Bryja J., Konvičková H., Russo I. R., Verheyen E. & Šumbera R. Phylogeography of a widespread sub-Saharan murid rodent *Aethomys chrysophilus*: the role of geographic barriers and paleoclimate in the Zambezian bioregion. *Mammalia*, 82, 4, 373-387, 2018.

Gryseels S, Goüy de Bellocq J, Makundi R, Vanmechelen K, Broeckhove J, **Mazoch V**, Šumbera R, Zima J, Leirs H, Baird SJE. Genetic distinction between contiguous urban and rural multimammate mice in Tanzania despite gene flow. *Journal of Evolutionary Biology*, 29, 10, 1952-1967, 2016.

McDonough MM, Šumbera R, **Mazoch V**, Ferguson AW, Phillips CD, Bryja J. Multilocus phylogeography of a widespread savanna–woodland adapted rodent reveals the influence of Pleistocene geomorphology and climate change in Africa's Zambezi region. *Molecular Ecology*, 24, 20, 5248-5266, 2015.

Bryja J, Mikula O, Šumbera R, Meheretu Y, Aghová T, Lavrenchenko LA, **Mazoch V**, Oguge N, Mbau JS, Welegerima, K. Pan-African phylogeny of *Mus* (subgenus *Nannomys*) reveals one of the most successful mammal radiations in Africa. *BMC Evolutionary Biology*, 14, 1, 256, 2014.

Bryja J, **Mazoch V**, Patzenhauerová H, Mateke C, Zima J, Šklíba J, Šumbera R. Revised occurrence of rodents from the tribe Praomyini (Muridae) in Zambia based on mitochondrial DNA analyses: implications for biogeography and conservation. *Folia Zoologica*, 61, 3-4, 268-283, 2012.

Mole rats

Kott O, Němec P, Fremlová A, **Mazoch V**, Šumbera R. Behavioural tests reveal severe visual deficits in the strictly subterranean African mole-rats (Bathyergidae) but efficient vision in the fossorial rodent Coruro (*Spalacopus cyanus*, Octodontidae). *Ethology*, 122, 8, 682-694, 2016.

Šumbera R, **Mazoch V**, Patzenhauerová H, Lövy M, Šklíba J, Bryja J, Burda H. Burrow architecture, family composition and habitat characteristics of the largest social African mole-rat: the giant mole-rat constructs really giant burrow systems. *Acta Theriologica*, 57, 2, 121-130, 2012.

Šklíba J, **Mazoch V**, Patzenhauerová H, Hrouzková E, Lövy M, Kott O, Šumbera R. A maze-lover's dream: Burrow architecture, natural history and habitat characteristics of Ansell's mole-rat (*Fukomys anselli*). *Mammalian Biology - Zeitschrift für Säugetierkunde*, 77, 6, 420-427, 2012.

Chimpanzees

Huffman MA, Petrželková K, Moscovice LR, Mapua MI, Bobáková L, **Mazoch V**, Singh J, Kaur T. Introduction of chimpanzees onto Rubondo Island National Park, Tanzania. *Global re-introduction perspectives: re-introduction case-studies from around the globe*, 213-215, 2008.

Petrášová J, Modrý D, Huffman M A, Mapua MI, Bobáková L, **Mazoch V**, Singh J, Kaur T, Petrželková KJ. Gastrointestinal parasites of indigenous and introduced primate species of Rubondo Island National Park, Tanzania. *International Journal of Primatology*, 31, 5, 920-936, 2010.

mazoch.vladimir@gmail.com

Diversity, evolution, and distribution of selected African rodents, with focus on East African savannahs.
Ph.D. Thesis Series, 2018, No. X

Printed in the Czech Republic by Typodesign
Edition of 20 copies

University of South Bohemia in České Budějovice
Faculty of Science
Branišovská 1760
CZ-37005 České Budějovice, Czech Republic

Phone: +420 387 776 201
www.prf.jcu.cz, e-mail: secret-fpr@prf.jcu.cz