Genetic variation of the most abundant forest-dwelling rodents in Central Africa (Praomys jacksoni complex) : evidence for Pleistocene refugia in both montane and lowland forests

Title: Genetic variation of the most abundant forest-dwelling rodents in Central Africa (*Praomys jacksoni* complex): evidence for Pleistocene refugia in both montane and lowland forests

Running title: Plio-Pleistocene evolution of *Praomys* rodents in African forests

Authors: Daniela Mizerovská [1,2,*], Violaine Nicolas [3,*], Terrence C. Demos [4,*], Dudu Akaibe [5], Marc Colyn [6], Christiane Denys [3], Prince K. Kaleme [7], Pionus Katuala [5], Jan Kennis [8], Julian C. Kerbis Peterhans [4,9], Anne Laudisoit [8,10], Alain Didier Missoup [11], Radim Šumbera [12], Erik Verheyen [8,13], Josef Bryja [1,2,**]

[1] Institute of Vertebrate Biology of the Czech Academy of Sciences, 603 65, Brno, Czech Republic
[2] Department of Botany and Zoology, Faculty of Science, Masaryk University, 611 37, Brno, Czech Republic
[3] Institut de Systématique, Evolution, Biodiversité (ISYEB), Muséum national d'Histoire naturelle, CNRS, Sorbonne Université, EPHE , CP51, 75005 Paris, France
[4] Integrative Research Center, Field Museum of Natural History, 60605, Chicago, IL, USA
[5] University of Kisangani, Animal Ecology and Resource Management Laboratory (LEGERA) and Biodiversity Monitoring Center (CSB), Kisangani, Democratic Republic of the Congo
[6] Université de Rennes 1, CNRS, UMR 6553 Ecobio, Station Biologique, 35380 Paimpont, France
[7] Laboratory of Mammalogy, Biology Department, CRSN Lwiro, Bukavu, DR Congo
[8] University of Antwerp, Evolutionary Ecology Group, Antwerp, Belgium
[9] College of Arts & Sciences, Roosevelt University, 60605, Chicago IL, USA
[10] EcoHealth Alliance, New York, USA
[11] Zoology Unit, Laboratory of Biology and Physiology of Animal Organisms, Faculty of Science, University of Douala, Douala, Cameroon
[12] Department of Zoology, Faculty of Science, University of South Bohemia, České Budějovice, Czech Republic
[13] Royal Belgian Institute for Natural Sciences, Operational Direction Taxonomy and Phylogeny, 1000, Brussels, Belgium

* equal contribution; other authors are ranked alphabetically

Correspondence: Josef Bryja, Institute of Vertebrate Biology of the Czech Academy of Sciences, Research Facility Studenec, Studenec 122, 675 02 Koněšín, Czech Republic; E-mail: bryja@brno.cas.cz

Acknowledgements

This study was supported by the Czech Science Foundation, project no. 14-36098G, by EU-DGVIII BIOFAC and ECOFAC programs, and by ANR Biodiversité IFORA Cameroon. For help in the field, logistics, and for providing samples, we acknowledge A. Konečný, J. Mbau, H. Konvičková, V. Mazoch, M. Lövy, J. Šklíba, J. Zima, S. Šafarčíková, members of UNIKIS, LEGERA and Congo2010 (S. G. Mbalitini,

C. K. Kyamakya, G. G. Tungaluna, P. M. Kabeya, N. A. Draso, P. Musaba), IRD Cameroon (Dr X. Garde, M. Elomo, C. Essono) and all local collaborators. A. Bryjová, H. Konvičková, and T. Aghová helped with genotyping in the Czech lab and O. Mikula with data analyses. The French part of the lab work was performed at the 'Service de Systématique Moléculaire' of the MNHN (UMS 2700, Paris, France). For permission to carry out the research and to specimen collection we are obliged to the Kenyan Forest Service and the Kenyan Wildlife Service (Kenya), Sokoine University of Agriculture in Morogoro (Tanzania), the Ministry of Research and Science (Yaoundé) in Cameroon, and the Zambian Wildlife Authority.

Submitted to: Journal of Biogeography as Research Paper

Abstract

Aim We investigate the Plio-Pleistocene evolutionary history of one of the most abundant rodents in Afrotropical forests. Specifically we ask how their diversification was influenced by climate change, topography and major rivers.

Location Tropical Africa: Lower Guinea (including Cameroon volcanic line; CVL), Congolia, Albertine Rift (AR), Kenyan highlands (KH).

Taxon Murine rodents of the *Praomys jacksoni* complex.

Methods We used 849 genotyped individuals to describe the overall diversity and spatial genetic structure across a majority of their known distribution area. The combination of one mitochondrial and three nuclear markers was used to infer dated phylogenies using Bayesian and maximum likelihood approaches. Genetic structure was further assessed by multi-species coalescent species delimitation. Current and past distributions of particular taxa were predicted by environmental niche modeling.

Results The complex is composed of five major genetic clades (proposed species). Two of them are restricted to specific habitat types (either montane forests of AR or wetlands in lowland forests along the Congo River), three others have wide geographic distributions and lower levels of ecological specialization. The earliest divergence is dated to the Plio-Pleistocene boundary and is in accordance with the separation of AR forests and Guineo-Congolian forests. Further diversification of the complex is associated with Pleistocene climate changes. Relatively stable refugia of suitable climatic conditions were identified in lowland Congolia (for two species currently distributed only in lowland forests) as well as in montane forests of CVL, AR, KH (playing the role of reservoirs of diversity). Large rivers, especially the Congo River, are important barriers to gene flow for most taxa, but probably were not the primary cause of differentiation.

Main conclusions The evolutionary history of the complex was primarily affected by Pleistocene climate changes and diversification in forest refugia. There is little support for ecological parapatric speciation or the riverine barrier hypothesis.

Keywords: Lowland forests, montane forests, phylogeography, Plio-Pleistocene climate changes, Praomyini, refugia, *Praomys jacksoni* species complex, Rodentia, tropical Africa.

Introduction

The African Guineo-Congolian rainforest biotic zone is the second largest block of tropical forests in the world after Amazonia (Barthlott, Lauer & Placke, 1996). Based mainly on the distribution of plants, these African rainforests can be divided into smaller biogeographical regions. Western (= Upper Guinean) forests are separated from West-Central (=Lower Guinean) forests by the Cross River and Cameroon Volcanic line (CVL) (Droissart et al., 2018; Fig. 1). Central African (= Congolian) forests are separated from Lower Guinean forests by the river Ubangi (Hardy et al., 2013) and can be divided into East-Central and South-Central forests by the river Congo. On the other hand, Eastern Afromontane forests are geographically much less extensive than Guineo-Congolian forests and occur mainly in relatively small patches along the East African Rift (Fig. 1). They are well known for very high proportions of endemic species (e.g. Plumptre et al., 2007), making them one of the most important biodiversity hotspots in the world. These montane forests can be divided into western (Albertine Rift mountains; ARM) and eastern (Kenyan Highlands = KH, Eastern Arc Mountains, Southern Rift Mountains) blocks. The ARM forests are adjacent to lowland Congolian forests and have very different vegetation (Fayolle et al., 2014, Droissart et al., 2018) but data describing the overlap of their fauna are scarce.

The current biodiversity of African forests was formed through the interplay of numerous factors including continental drift, geological activity, past climatic changes, and biotic factors like stochastic dispersal events or interspecific interactions (Plana, 2004). Guineo-Congolian, Eastern Afromontane, and Eastern coastal forests, all currently clearly separated, were probably linked in a continuous belt during the warm and humid Early to Middle Miocene (23-5 Ma) (Plana, 2004). At the end of the Miocene (8 Ma), the proportion of C4 biomass in tropical Africa increased (Cerling, 1992), which is considered to be due to expansion of savanna grasses and partial replacement of lowland rainforest by savanna woodland (Jacobs, 2004). The Eastern Arc Mountains (and parts of ARM) served as long-term refugia for eastern Afromontane forests (Plana, 2004). The Pliocene (from 5.3 Ma onwards), and especially the Pleistocene (starting at 2.5 Ma), are known as periods with dramatic oscillations between drier and more humid conditions. Several periods of highly variable climate (Potts, 2013; deMenocal, 2004) likely caused fragmentation of rainforests into refugia leading to the allopatric diversification of forest-dwelling African fauna.

Among mammals, rodents are good candidates for describing and understanding the evolutionary history of particular ecosystems. They have short generation times, rapid mtDNA substitution rates, strong associations with specific habitats, and limited dispersal ability. Recently, African rodents have been used as biogeographical models for reconstruction of the evolutionary history of savannas (e.g. Aghová et al., 2017, Mazoch et al., 2018), as well as various types of forests (e.g. Bryja, Mikula, Patzenhauerová et al., 2014; Bohoussou et al., 2015). For assessing the history and biogeography of Afrotropical forests, the murine tribe Praomyini is an appropriate model, because members of this tribe are forest specialists with abundant populations, and their phylogenetic history may mirror the

history of their habitats (e.g. Nicolas et al., 2008; Demos, Kerbis Peterhans, Agwanda & Hickerson, 2014).

The genus *Praomys* (Thomas, 1915) has ca. 20 recognized species divided into five species complexes (*lukolelae*, *daltoni*, *jacksoni*, *tullbergi*, *delectorum*) distributed in Afrotropical forests and the adjacent forest-savanna mosaic (Denys, Taylor & Aplin, 2017). The evolutionary history and phylogenetic relationships in three of the five *Praomys* complexes were recently resolved by the application of molecular genetic analyses of DNA sequences (the *daltoni* complex, Bryja et al., 2010; the *tullbergi* complex, Missoup et al., 2012 and references therein; and the *delectorum* complex, Bryja, Mikula, Patzenhauerová et al., 2014). Multiple phylogeographic studies, especially in Upper and Lower Guinean forests, found this genus very suitable for testing hypotheses of diversification in tropical forests (sensu Haffer, 1997) including the refuge hypothesis, the riverine barrier hypothesis, and the hypothesis of ecological gradients (Moritz, Patton, Schneider & Smith, 2000). For example, Nicolas et al. (2008) used two sibling species in the *tullbergi* complex, *P. tullbergi* (Thomas, 1894) and *P. rostratus* Miller, 1900, distributed in partial sympatry in Upper Guinean forests in West Africa, to test the role of habitat specialization level on their genetic architecture. The most widespread species from the same group, *P. misonnei* Van der Straeten & Dieterlen, 1987, was used as a biogeographical model to test the relative role of rivers and lowland forest refugia as drivers of diversification (Nicolas et al., 2011). Finally, the diversification patterns, especially in relation to different elevations in west-central Africa, were analysed in all species of the *tullbergi* complex (Missoup et al., 2012).

Despite the fact that members of the *P. jacksoni* complex are widespread in Lower Guinea, Congolia, and part of Eastern Afromontane hotspot, and are often the most abundant members of rodent assemblages, genetic studies of this group have been geographically very restricted (e.g. Kennis et al., 2011, Kisangani region in Democratic Republic of Congo (=DRC); Bryja et al., 2012, Zambia). Furthermore, all these studies relied on mitochondrial sequence data alone; thus, the complete picture of biogeographical patterns remains obscure. The absence of large-scale genetic data limits inferences about the evolutionary history of the *P. jacksoni* complex and the historical biogeography of central African rainforests. Our study is the first to use multi-locus genetic data to analyse the geographic distribution of the genetic variability within the entire *P. jacksoni* complex. Using the most extensive available collection of tissues from all species in the complex collected across most of their distribution (ca. 850 specimens), phylogenetic reconstructions were carried out in a temporal framework and bioclimatic niches (i.e. extent of climatically suitable habitats) of particular taxa were modelled in current conditions as well as during the last glacial cycle. Finally, we discuss how geomorphology and Plio-Pleistocene climate changes might have affected the evolutionary history of these forest specialist mammals.

Materials and methods

Sampling and genotyping

Over the past 30 years, animals were prepared as vouchers and tissue samples (spleen, kidney, muscle or toe) were stored in 96% ethanol, saturated salt solution, or liquid nitrogen. The members of the *P. jacksoni* complex are generally the most abundant component of small mammal communities and none is listed as endangered. All fieldwork complied with environmental regulations in the respective African countries and sampling was carried out in accordance with local legislation (see Acknowledgements). Data from 597 newly genotyped specimens were supplemented by previously analysed material, whose sequences were available in GenBank (252 individuals). In total, we assembled genetic data from 849 individuals from 86 localities in 11 countries (Fig. 1, Appendix S1; Fig. S2.1 in Appendix S2).

Four genetic markers were used for analyses of genetic structure and phylogenetic inference. All individuals were genotyped for the mitochondrial gene cytochrome *b* (*CYTB*), using the protocol of Bryja, Mikula, Šumbera et al. (2014). Selected specimens from each major mitochondrial clade (see Appendix S1) were also genotyped at three unlinked autosomal introns: *ACPT*, *CARHSP*, and *GAD2* (see details in Demos et al., 2014). Genetic markers were amplified by polymerase chain reaction (PCR) and commercially sequenced by the Sanger method.

Mitochondrial phylogeny and genetic distances

The number of genetic partitions in the *CYTB* alignment and the most suitable nucleotide substitution models were simultaneously estimated in PartitionFinder 2 (Lanfear, Frandsen, Wright, Senfeld & Calcott, 2016). The best supported models were used for partitioned phylogenetic reconstructions by Bayesian inference (BI) in MrBayes 3.2.6 (Ronquist & Huelsenbeck, 2003) and the maximum likelihood (ML) approach in RAxML 8.0 (Stamatakis, 2014). More details on phylogenetic analysis of mtDNA sequences are provided in Appendix S3 (Supplementary Material).

Genetic distances among mtDNA clades were calculated in MEGA 6.06 (Tamura, Stecher, Peterson, Filipski & Kumar, 2013) as *p*-distances and Kimura 2-parameter (K2P) distances. Two approaches were used to examine the geographic distribution of genetic variation within major mitochondrial clades. First, 192 *CYTB* sequences from *P. jacksoni* sensu lato (see Results) were trimmed to 714 bp, haplotypes were identified in DnaSP 5.10.01 (Librado & Rozas, 2009) and a haplotype network calculated by the median-joining method in Network 5.0.0.1 (Bandelt, Forster & Röhl, 1999). Second, using simple dispersal scenarios, we estimated the location of ancestral populations as coincident with the geographic region of maximum genetic diversity (Excoffier, Foll & Petit, 2009) using an algorithm called ”genetic hubs” (Mikula, 2018). The algorithm can be explored using the package 'GenHubs' for R 3.3.1 (R Core Team, 2016), which is provided in Appendix S3.

Coalescent species delimitation

Our results indicated 11 mtDNA lineages that may represent evolutionarily isolated gene pools (=species). We tested their distinctiveness using the combined mitochondrial and nuclear gene dataset and the nuclear genes dataset alone in a Bayesian framework using BP&P 3 software (Yang &

Rannala, 2014). A speciation probability of 1.0 on a node indicates that every species delimitation model visited by the rjMCMC algorithm supports the hypothesis that the two lineages descending from a particular node represent independently evolving gene pools. We evaluated the influence of priors on the posterior probability distribution by testing priors for θ and τ0, assuming either small or large ancestral population size and shallow or deep divergences (see details in Appendix S3).

Species tree and divergence dating

To analyse the relationships among species delimited in BP&P in a temporal context, we estimated divergence times in a species tree using the multispecies coalescent model as implemented in StarBEAST 2 (Ogilvie, Bouckaert & Drummond, 2017). For this analysis all four loci sequenced in 55 individuals, representing 11 taxa delimited by BP&P (see Results), with 3-7 individuals per species were used (Appendix S1). Because *Praomys* has a poor fossil record, it is not possible to calibrate the molecular clock by ingroup fossils. We therefore performed a secondary calibration using the time to most recent common ancestor (TMRCA) of the *P. jacksoni* complex estimated by Aghová et al. (2018). More details about specification of priors and evaluation of outputs can be found in Appendix S3.

Ecological niche modelling

The geographic distributions of five putative species in the *P. jacksoni* complex were estimated using the MaxEnt algorithm (Phillips, Anderson & Schapire, 2006) for the last interglacial (LIG; 120-140 ka), the last glacial maximum (LGM; 22 ka), and present conditions. The purpose of this analysis was to establish the spatial and temporal framework for potential geographic connections among sampled populations. As predictors, we used 19 bioclimatic variables downloaded from the WorldClim database (Hijmans, Cameron, Parra, Jones & Jarvis, 2005) and downsampled to 0.5° resolution. Response was represented by unique presence records left after mapping of georeferenced barcoded individuals to 0.5° grid. More details about the models and interpretation of results are in Appendix S3.

Results

Mitochondrial phylogeny and distribution of genetic diversity

In total, we obtained 849 sequences of mitochondrial *CYTB* belonging to the *P. jacksoni* complex. For inference of the mitochondrial gene tree we used 221 unique sequences (Appendix S1), belonging to the main clades identified in preliminary analysis. The remaining sequences were unambiguously assigned to particular mtDNA lineages by neighbour-joining analysis in MEGA (bootstrap support > 90%; not shown). These data were used mainly to increase the precision by which the geographical distribution of phylogenetic clades was mapped.

Both BI and ML phylogenetic analyses provided similar topologies of the mtDNA tree (Fig. 2 and Fig. S2.2 in Appendix S2) and confirmed the monophyly of the *P. jacksoni* complex. The complex is further subdivided into three well-supported major clades that can be assigned to nominal species *P.*

degraaffi, *P. mutoni* and *P. jacksoni* sensu lato, with the two latter taxa being significantly supported sisters. *Praomys jacksoni* sensu lato is composed of four monophyletic clades (I-IV); one of them, clade III, corresponds to the named species *P. minor* (sensu Kennis et al. 2011), but it is not sister to the remaining three clades (Fig. 2), making *P. jacksoni* (clades I+II+IV) paraphyletic.

Praomys degraaffi is a montane species, endemic to the central ARM (Fig. 1a). *Praomys mutoni* was recorded only in lowland forests in DRC on both banks of the middle and upper Congo River, albeit with the easternmost locality within the foothills of ARM (Fig. 1b). *Praomys minor* (= *P. jacksoni* sensu lato clade III) had a similar distribution as *P. mutoni*, but was found only on the left bank of the Congo River, up to north-western Zambia (Fig. 1a). The sequences of *P. minor* from Zambia formed a distinct haplogroup, compared to those from DRC (Fig. 3).

In addition to *P. minor*, *P. jacksoni* sensu lato includes three additional clades (I, II, IV) with parapatric distributions. The most distinct is clade IV (Fig. 2), where we recognize three haplogroups based on the haplotype network (Fig. 3). Clade IVa is distributed in Lower Guinea, including both the lowland and montane forests (Fig. 1c). The haplotypes from Mt. Oku and Mt. Lefo in CVL are clearly separated from the rest of the haplogroup (Fig. 3). Clade IVb is widely distributed in the central part of the Congo Basin (on both banks of the Congo River), reaching the southernmost part of Central African Republic (CAR). Finally, clade IVc is present in one locality at the right bank of the Congo River in DRC and in north-central CAR (Fig. 1c).

Sister clades I and II (= *P. jacksoni* sensu stricto) from the central-eastern African forest are parapatric to the clade IV, with which they might overlap in central Congolia (Figs. 1c vs. 1d). They also have internal structure (Figs. 2 and 3). In clade I we recognize three haplogroups: Ia has a very limited distribution on Mount Kenya, Ib is widespread mainly in the forests along the ARM with geographically distant populations from the Kisangani region (Bomane) forming a separate subgroup of haplotypes (Fig. 3), and Ic includes specimens from both Kenyan lowlands (Kakamega forest) and highlands, excluding Mount Kenya. In KH, a distinct subgroup from Mount Elgon is apparent in the network of the Ic haplogroup (Fig. 3). Clade II has two haplogroups (Fig. 3): IIa was recorded in northern Zambia and easternmost DRC (with one record in Burundi and one in Mbizi forest in Tanzania), while IIb was found only on the right bank of the Congo River in the Kisangani region (Fig. 1d).

The genetic hubs algorithm identified the regions with the highest mitochondrial diversity, i.e. potential long-term refugia in suitable habitats (Fig. 4). For *P. degraaffi* the hub is localized in the central part of ARM (Virunga Mts.), while *P. mutoni* has the highest diversity in the Yoko region on the left bank of the Congo River. Subclades of *P. jacksoni* clade IV had the highest diversity in the central part of CVL (IVa), Congo-CAR border (IVb) and central CAR (IVc). The latter can be biased by unequal sampling (only two discontinuous areas) and the same is true for *P. minor*, where the hub

was predicted in Zambia. Remaining clades of *P. jacksoni* have genetic hubs in ARM (Ib, IIa), KH (Ia on Mt. Kenya and Ic on Mt. Elgon), and in Kisangani region in DRC (IIb).

Genetic distances (both *p*-distances and K2P-distances) between individual clades calculated from *CYTB* data are shown in Tables S2.1 and S2.2 in Appendix S2. Distances among 11 mitochondrial clades ranged from 0.0234 to 0.1004, with all values < 0.05 representing pairwise distances between individual subclades of *P. jacksoni* clade IV and *P. jacksoni* sensu stricto (comprising clades I and II). All distances between the five major clades (i.e. *P. degraaffi*, *P. mutoni*, *P. jacksoni* clade IV, *P. minor* and *P. jacksoni* sensu stricto) were > 0.05.

Species delimitation

All BP&P analyses provided highly significant ESS values (>200) regardless of the dataset (only nuclear markers vs. combined nuclear and mitochondrial data), priors (i.e. various combination of τ0 and θ), and algorithms (0 or 1) (Appendix S4). All analyses significantly supported *P. degraaffi*, *P. mutoni*, and *P. minor* as distinct species (PP = 1). In *P. jacksoni* sensu lato, clades Ia, Ib, Ic, IIa and IIb were identified as distinct gene pools (PP > 0.97) in all analyses. The split of clade IV into several gene pools is not clear. When using nuclear loci alone, the populations from mitochondrial clades IVa + IVb were grouped together (PP = 0.58-0.66), but the clade IVc was supported as a separate species (PP = 1). When *CYTB* sequences were included in the dataset, all three mitochondrial clades were recognized as distinct species with PP = 0.92-0.95 for clades IVa and IVb, and PP = 1 for clade IVc.

Dating of divergence

The species tree based on the combined mitochondrial and nuclear datasets (Fig. 5a) has a topology similar to the mitochondrial tree. It differs mainly in the positions of *P. mutoni* and *P. degraaffi*, with the former being the sister to all remaining taxa in the species tree, albeit with relatively low posterior probability, and by a branching pattern within *P. jacksoni* sensu stricto. The first split is estimated to 3.0 Ma, and the MRCA of *P. degraaffi* and *P. jacksoni* sensu lato is dated at 2.2 Ma. Spatial structure within *P. jacksoni* clade IV and *P. jacksoni* sensu stricto is estimated to have arisen in the late Pleistocene (< 0.7 Ma). When we performed the analysis using only nuclear markers (Fig. 4b), *P. minor* appeared as very distinct taxon, diverging after *P. mutoni* at the beginning of Pleistocene (2.1 Ma). Interestingly, in both analyses that involved nuclear genes, *P. jacksoni* clade Ib forms a monophyletic group with other populations of *P. jacksoni* sensu stricto from ARM and Congolia (i.e. IIa + IIb), separated from KH populations (Ia + Ic). The results were very similar when we considered only five species representing major mitochondrial clades instead of 11 taxa identified by BP&P analysis (Fig. S2.3 in Appendix S2).

Ecological niche modelling

The MaxEnt models were analysed separately for each of five major clades (Fig. 6), because they likely represent taxa with different ecological requirements. The AUC values indicate good model performance for all five taxa (AUC ranging from 0.93 to 0.98). Predicted distributions in the present

are slightly larger than known occurrence evidenced from genotyped material for all but one species (*P. degraaffi*). This is not surprising given the relatively poor sampling in South-Central Congolia. All species, except *P. degraaffi*, are predicted to be widely distributed in Central Africa, but the probabilities of presence in a given region differ from one species to another. *Praomys jacksoni* sensu stricto has a higher probability of presence near ARM, while *P. jacksoni* clade IV has a higher probability of presence in Lower Guinea. The probabilities of presence of *P. minor* and *P. mutoni* are especially high in South-Central Congolia. *Praomys degraaffi* is a highly specialized species with narrow bioclimatic requirements, and the model predicted its distribution only in a very limited range in ARM. Unexpectedly, the models predict similar distributions of most taxa at LGM compared to the present, i.e. the model does not support the presence of geographically restricted climatic refugia, at least at the LGM. The predicted distributions for LIG are generally smaller, and this is particularly apparent for the lowland species *P. minor* and *P. mutoni*. Areas of climatic stability across the last glacial cycle for remaining species are localised in mountain areas (CVL for *P. jacksoni* clade IV, KH and ARM for *P. jacksoni* sensu stricto, and ARM for *P. degraaffi*).

Discussion

Distribution of the complex - is it a suitable model for biogeographic reconstruction of Afrotropical forests?

In this study, we present the most comprehensive molecular phylogeny of the *P. jacksoni* complex to date. The strongly supported monophyletic *P. jacksoni* complex is one of three major clades unambiguously included in the genus *Praomys* (if we omit the *P. lukolelae* complex with unclear phylogenetic relationships and the *P. delectorum* clade that should be excluded from the genus based on genetic data; Missoup et al., 2012). The three *Praomys* complexes differ in their biogeographic patterns. The *P. daltoni* complex is distributed in the mosaic of the Guinean forest and Sudanian savanna (Bryja et al., 2010), while the *P. tullbergi* complex has the highest diversity in Lower Guinean forests, especially in CVL (Missoup et al., 2012). In contrast, we show that the *P. jacksoni* complex has its highest diversity in the Congolian forests and ARM, with a single clade extending into Lower Guinea.

Individual taxa within the *P. jacksoni* complex differ in their habitat requirements. *Praomys degraaffi* is a montane forest specialist restricted to ARM at elevations above 1500 m a.s.l. (e.g. Van der Straeten & Kerbis Peterhans, 1999; Kaleme, 2011). *Praomys mutoni* is a typical lowland species, the distribution of which has been limited to relatively small area in the Kisangani region (DRC), where it lives in swampy areas and riverine forests on both banks of the Congo River (Nicolas et al., 2005; Katuala et al., 2008; Kennis et al., 2011). We reanalysed two specimens from the locality Bushema Lutunguru (reported by Kaleme, 2011) and added two new localities between the Congo and Lomani Rivers, which almost doubles its known distribution (Denys et al., 2017). We also modelled the distribution of these two species using bioclimatic data. The predicted distribution of *P. degraaffi* remained limited to ARM, but *P. mutoni* might have a wider distribution in the humid lowland forests of the central Congo basin, i.e. a region that is still largely unsurveyed.

The remaining taxa have less strict ecological requirements. *Praomys jacksoni* clade IV is distributed mainly in Lower Guinea and is a generalist taxon whose habitats include both montane (in CVL) and lowland rainforest. In Lower Guinea it was only captured in small forest patches or gallery forest embedded in savanna (V. Nicolas, pers. obs.). In the lowland Kisangani region at its range limit, it has even been collected in human-degraded habitats such as fallow palm plantations and regenerating agricultural plots (Kennis et al., 2011). *Praomys minor* is a typical lowland species distributed in primary and secondary forest on the left bank of the Congo River. Kennis et al. (2011) also reported this species in degraded habitats, e.g. fallow land and plantations. On the other hand, in Zambia it was found only in gallery forest and evergreen moist forest near the Zambezi source (Bryja et al., 2012). Finally, *P. jacksoni* sensu stricto (clades I + II) is present in multiple habitats, including lowland primary and secondary forests, fallow lands, and montane forests of ARM and KH (e.g. Katuala et al., 2008; Kaleme, 2011; Kennis et al., 2011). However, even in the most degraded habitats at least some tree cover is always required (e.g. small riverine forests in otherwise open landscape as observed in northwestern Tanzania; J. Bryja, pers. obs.). The particular clades of the complex can occupy different ecological niches, but they always require tropical forests (or ecotones). Analyses of their genetic structure can thus provide information needed to infer the evolutionary history of forests in Lower Guinea, Congolia, ARM, and KH.

Reconstruction of evolutionary history - the role of climatic changes, mountains and rivers

The time of the first divergence in the genus *Praomys* producing three unambiguously supported species complexes is estimated at ca. 5 Ma and the medians of first splits within these complexes are estimated as the late Pliocene/early Pleistocene: 2.2-3.3 Ma for the *P. jacksoni* complex (Lecompte, Granjon, Kerbis Peterhans & Denys, 2002; this study), 2.5-3.3 Ma for the *P. tullbergi* complex (Missoup et al., 2012), and 3.0 Ma for the *P. daltoni* complex (Bryja et al., 2010). Although it is difficult to compare different studies because of their differing molecular clock calibrations, it is evident that most speciation events in all three *Praomys* clades occurred in the Pleistocene and may have been affected by climatic changes in last ca. 2.5 Ma.

The *P. jacksoni* complex very likely started to diversify in Central Africa, where we observe the highest genetic diversity. The first cladogenetic split is not well resolved. Based on the mtDNA tree *P. degraaffi* is sister to all remaining taxa, while combined nuclear + mtDNA data and nuclear data alone support *P. mutoni* as sister to the remaining *Praomys* clades. In either case, the first divergence is coincident with the isolation of montane forests in the Albertine Rift from Congo Basin lowland forests due to increasing aridification at the Pliocene/Pleistocene boundary (Plana, 2004). The evolutionary processes that affected further diversification can be assessed from the distribution of genetic variability today. For example, the distribution of *P. mutoni* exclusively in lowland forests supports a fluvial refuge model, which stipulates that gallery forests acted as long-term refugia during glacial cycles. In Africa, the hypothesis that lowland forest patches persisted near rivers in the central Congo Basin is supported by the distribution of diversity of primates and plants (e.g., Colyn,

Gautier-Hion & Verheyen, 1991; Robbrecht, 1996). New data from the geographically large, and to date undersampled South-Central Congolia is required to test this hypothesis; for example by comparison of genetic structure of *P. mutoni* and *P. minor* with fluvial networks.

Mountains on both sides of the Congo Basin (i.e. CVL and ARM) have very likely played an important role in the evolution of taxa now distributed in Lower Guinea and East-Central Congolia. Both mountain ranges are inhabited by the same taxa of the *P. jacksoni* complex as neighbouring lowland forests (except the montane ARM specialist *P. degraaffi*) and for multiple subclades the highest genetic diversities ("genetic hubs") were found in the mountains. This suggests that long-term refugia for taxa distributed on the right bank of the Congo River may have been present in mountains, which thus served as "museums" or "sinks" of diversity (Stebbins, 1974; Plana, 2004). From montane refugia (both east and west), populations may have dispersed across the lowland rainforests of central Africa (see genetic hubs for *P. jacksoni* clade Ib and IVa in Fig. 4). This hypothesis is reinforced by the climatic niche modelling analyses showing that for *P. jacksoni* sensu stricto and *P. jacksoni* clade IV, areas of climatic stability throughout the Pleistocene are localised in ARM and CVL, respectively. Very similar phylogeographic structure has been recently documented for two other forest rodents widely distributed on the right bank of the Congo River, *P. misonnei* (Nicolas et al., 2011) and *Malacomys longipes* (Bohoussou et al., 2015). One major clade occurs in the west (including CVL) and one in the east (including ARM), which may indicate the generally important role of forest refugia in mountains neighbouring the Congo basin even for taxa currently distributed in lowland forests. The hypothesized role of mountains as reservoirs of diversity in the *P. jacksoni* complex contrasts with inferred speciation patterns in the *P. tullbergi* group. The phylogenetic analysis of Missoup et al. (2012) suggests that highland species in montane Cameroonian forests likely evolved by parapatric speciation along an elevational gradient from lowland taxa, where CVL mountains may have acted as speciation "engines" (Plana, 2004).

The major biogeographic divide within the *P. jacksoni* complex is clearly the Congo River (Kennis et al., 2011), but it seems unlikely that it acted as a primary driver of diversification (i.e. the "riverine barrier hypothesis", which posits that a widespread ancestral population was split when large rivers developed; Haffer, 1997). Instead, the Congo River may have blocked the range expansion of those taxa that had already speciated in allopatry in isolated forest refugia. For example, *P. minor* is probably a widespread taxon in South-Central Congolia, but its distribution is abruptly terminated by the Congo River in the north. In contrast, *P. jacksoni* sensu stricto is only found on the right bank of the Congo River in the Kisangani region (Fig. 1). Very similar patterns have been documented in other forest rodents (Nicolas et al., 2011; Bohoussou et al., 2015) and primates (Eriksson et al., 2004) in the Congo Basin. As previously reported, dispersal across the Congo River has occurred at least twice by the members of the *P. jacksoni* complex (Kennis et al., 2011). First, the river is not a barrier for *P. mutoni*, a rainforest swamp specialist, and probably an adept swimmer. More surprisingly, *P. jacksoni* clade IV was also found on both banks of the Congo River in the Kisangani region (Fig. 1), but in this case the two populations differ genetically. The population on the left bank (clade IVb) is genetically

similar to populations found at the Republic of Congo-CAR border (right bank), where a "genetic hub" of this clade was located. It is possible that the eastward expansion of this lineage across the Congo River occurred at the DRC and Republic of Congo border, where the river can be crossed more easily (Kennis et al., 2011).

Similarly to other mammals (Demos et al., 2014, 2015; Bryja et al., 2017) and plants (Plumptre et al., 2007; Droissart et al., 2018), our genetic analysis supports biogeographic affinities between ARM and KH. These two mountain massifs are currently separated by a 500 km wide gap without forest ('Uganda gap'), which currently forms a filter corridor for small mammals restricted to humid montane habitats (Demos et al., 2014). Recent phylogeographic and population genetic studies of montane forest-dwelling mammals (*Hylomyscus denniae* group and *Sylvisorex granti* group; Demos et al., 2014, 2015) and their comparison with less specialized *P. jacksoni* sensu stricto allow assessment of the effect of habitat specialization on phylogeographic structure. First, *P. jacksoni* sensu stricto is much more widespread than montane forest specialists are. Weaker ecological specialization allowed its dispersal from highland refugia (ARM and KH) into numerous lowland forests (Fig. 1). Second, the level of genetic structure in KH is higher in *P. jacksoni* than in more specialized forest taxa. The level of divergence between the two KH clades, Ia and Ic, is similar to that between Ib (in ARM) and KH. Demos et al. (2015) assumed that many local populations of *Hylomyscus* and *Sylvisorex* in KH went extinct during unsuitable Pleistocene periods and current forests were recolonized from a small number of founders, making them genetically homogenous. In contrast, *P. jacksoni* is not as strongly forest-restricted, which could have allowed persistence in more KH refugia (e.g. Mt. Kenya, Mt. Elgon, Aberdare Mts.). Thirdly, the split of KH and ARM lineages in both *Sylvisorex* and *Hylomyscus* is dated to at least the beginning of Pleistocene, ca. 2 Ma (Demos et al., 2015), which is reflected in greater divergence between ARM and KH and separate species status for KH and ARM lineages in both genera. In comparison, *P. jacksoni* in KH and ARM diverged ca. 0.5 Ma. This is again in agreement with the lower ecological specialization of this taxon, which could facilitate more recent gene flow across the Ugandan gap during the Pleistocene.

Taxonomic implications

The comprehensive multi-locus genetic dataset was also used to delimit species in the complex and the multi-species coalescent supported up to 11 separate gene pools. However, this approach has recently been shown to diagnose genetic structure, with no distinction between structure due to population isolation or due to speciation (Sukumaran & Knowles, 2017). It is therefore necessary to evaluate the results with caution, particularly as it serves as the sole basis for taxonomic revision. The multi-locus phylogeny supported five major clades (*P. mutoni*, *P. degraaffi*, *P. minor*, *P. jacksoni* clade IV and *P. jacksoni* sensu stricto), and we hypothesize that these may represent separate species. Some of them are relatively well characterized by ecology and morphology (*P. degraaffi*, *P. mutoni*, *P. minor*; Kennis et al., 2011) and there is little doubt that they are distinct biological entities. On the other hand, the species status of two remaining major clades in this study are more ambiguous and require future study. There are several lines of evidence that support *P. jacksoni* clade IV as a valid

species, distinct from *P. jacksoni* sensu stricto. First, it differs from other clades at *CYTB* by 6.75-9.32%, which is well in the range of other interspecific distances in *Praomys* (e.g. Missoup et al., 2012; this study). While acknowledging that the use of mtDNA can lead to biases in species delimitation, genetic distances at mtDNA are useful and simple tools to indicate species limits in rodents. The level of genetic differentiation at *CYTB* (K2P- or *p*-distance) between closely related sister species is generally near 5% in the tribe Praomyini (Lecompte et al., 2002). Second, even with relatively limited data for the nuclear markers, we found three fixed diagnostic SNPs separating individuals from mitochondrial clades I and IV (one in GAD2 and two in CARHSP genes) in their parapatric contact zone in the Kisangani region, on the right bank of the Congo River. More detailed integrative taxonomic study is required, but our data indicate that there is no (or very limited) gene flow between these parapatric taxa in their contact zone, supporting the hypothesis of their reproductive isolation.

References

check against format in guidance for authors.

Aghová, T., Kimura, Y., Bryja, J., Dobigny, G., Granjon, L. & Kergoat, G. J. (2018). Fossils know it best: using a new set of fossil calibrations to improve the temporal phylogenetic framework of murid rodents (Rodentia: Muridae). *Molecular Phylogenetics and Evolution*, 128, 98-111.

Aghová, T., Šumbera, R., Piálek, L., Mikula, O., McDonough, M. M., Lavrenchenko, L. A., ... Bryja, J. (2017). Multilocus phylogeny of East African gerbils (Rodentia, *Gerbilliscus*) illuminates the history of the Somali-Masai savanna. *Journal of Biogeography*, 44, 2295–2307.

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

Barthlott, W., Lauer, W. & Placke, A. (1996). Global distribution of species diversity in vascular plants: towards a world map of phytodiversity. *Erdkunde*, 50, 317-327.

Bohoussou, K. H., Cornette, R., Akpatou, B., Colyn, M., Kerbis Peterhans, J. C., Kennis, J., … Nicolas, V. (2015). The phylogeography of the rodent genus *Malacomys* suggests multiple Afrotropical Pleistocene lowland forest refugia. *Journal of Biogeography*, 42, 2049–2061.

Bryja, J., Granjon, L., Dobigny, G., Patzenhauerová, H., Konečný, A., Duplantier, J. M., … 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. *Molecular Ecology*, 19, 4783–4799.

Bryja, J., Mazoch, V., Patzenhauerová, H., Mateke, C., Zima, J., 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, 268–283.

Bryja, J., Mikula, O., Patzenhauerová, H., Oguge, N. O., Šumbera, R. & Verheyen, E. (2014). 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., Mikula, O., Šumbera, R., Meheretu, Y., Aghová, T., Lavrenchenko, L. A., … Verheyen, E. (2014). Pan-African phylogeny of *Mus* (subgenus *Nannomys*) reveals one of the most successful mammal radiations in Africa. *BMC Evolutionary Biology*, 14, 256.

Bryja, J., Šumbera, R., Kerbis Petterhans J. C., Aghová, T., Bryjová, A., Mikula, O., … Verheyen, E. (2017). Evolutionary history of the thicket rats (genus *Grammomys*) mirrors the evolution of African forests since late Miocene. *Journal of Biogeography*, 44, 182–194.

Cerling, T.E. (1992) Development of grasslands and savannas in East Africa during the Neogene. *Palaeogeography, Palaeoclimatology, Palaeoecology*, 97, 241–247.

Colyn, M., Gautier-Hion, A. & Verheyen, W. (1991). A re-appraisal of palaeoenvironmental history in Central Africa: evidence for a major fluvial refuge in the Zaire Basin. *Journal of Biogeography*, 18, 403–407.

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

Demos, T. C., Kerbis Peterhans, J. C., Agwanda, B. & Hickerson, M. J. (2014). Uncovering cryptic diversity and refugial persistence among small mammal lineages across the Eastern Afromontane biodiversity hotspot. *Molecular Phylogenetics and Evolution*, 71, 41–54.

Demos, T. C., Kerbis Peterhans, J. C., Joseph, T. A., Robinson, J. D., Agwanda, B. & Hickerson, M. J. (2015). Comparative population genomics of African montane forest mammals support population persistence across a climatic gradient and Quaternary climatic cycles. *PLoS ONE*, 10, e0131800.

Denys, C., Taylor, P. J. & Aplin, K. P. (2017). Family Muridae. In D. E. Wilson, T. E. Jr. Lacher & R. A. Mittermeier (Eds.), *Handbook of the Mammals of the World, Volume 7: Rodents II*. Barcelona: Lynx Edicions.

Droissart, V., Dauby, G., Hardy, O. J., Deblauwe, V., Harris, D. J., Janssens, S., … Couvreur, T. L. P. (2018). Beyond trees: Biogeographical regionalization of tropical Africa. *Journal of Biogeography*, 45, 1153–1167.

Eriksson, J., Hohmann, G., Boesch, C. & Vigilant, L. (2004). Rivers influence the population genetic structure of bonobos (*Pan paniscus*). *Molecular Ecology*, 13, 3425–3435.

Excoffier, L., Foll, M. & Petit, R. J. (2009). Genetic consequences of range expansions. *Annual Review of Ecology, Evolution, and Systematics*, 40, 481–501.

Fayolle, A., Swaine, M. D., Bastin, J. F., Bourland, N., Comiskey, J. A., Dauby, G., … Plumptre, A. J. (2014). Patterns of tree species composition across tropical African forests. *Journal of Biogeography*, 41, 2320–2331.

Haffer, J. (1997). Alternative models of vertebrate speciation in Amazonia: an overview. *Biodiversity and Conservation*, 6, 451–476.

Hardy, O. J., Born, C., Budde, K., Daïnou, K., Dauby, G., Duminil, J., … Poncet, V. (2013). Comparative phylogeography of African rain forest trees: A review of genetic signatures of vegetation history in the Guineo-Congolian region. *Comptes Rendus Geoscience*, 345, 284–296.

Hijmans, R. J., Cameron, S. E., Parra, J. L., Jones, P. G. & Jarvis, A. (2005). Very high resolution interpolated climate surfaces for global land areas. *International Journal of Climatology*, 25, 1965–1978.

Jacobs, B. F. (2004). Palaeobotanical studies from tropical Africa: relevance to the evolution of forest, woodland and savannah biomes. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 359, 1573–1583.

Kaleme, P. K. (2011). *Habitat fragmentation, patterns of diversity and phylogeography of small mammal species in the Albertine rift*. PhD thesis, Stellenbosch University, South Africa.

Katuala, P. G. B., Kennis, J., Nicolas, V., Wendelen, W., Hulselmans, J., Verheyen, E., … Leirs, H. (2008). The presence of *Praomys*, *Lophuromys*, and *Deomys* species (Muridae, Mammalia) in the forest blocks separated by the Congo River and its tributaries (Kisangani region, Democratic Republic of Congo). *Mammalia*, 72, 223–228.

Kennis, J., Nicolas, V., Hulselmans, J., Katuala, P. G. B., Wendelen, W., Verheyen, E., … Leirs, H. (2011). The impact of the Congo River and its tributaries on the rodent genus *Praomys*: speciation origin or range expansion limit? *Zoological Journal of the Linnean Society*, 163, 983–1002.

Lanfear, R., Frandsen, P. B., Wright, A. M., Senfeld, T. & Calcott, B. (2016). PartitionFinder 2: new methods for selecting partitioned models of evolution for molecular and morphological phylogenetic analyses. *Molecular Biology and Evolution*, 34, 772–773.

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? *Comptes Rendus Biologies*, 325, 827–840.

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

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, 373–387.

Mikula, O. (2018). Genetic hubs: a phylogeographer's widget for pinpointing of ancestral populations. BioRxiv preprint first posted online Sep. 18, 2018; doi: http://dx.doi.org/10.1101/419796.

Missoup, A. D., Nicolas, V., Wendelen, W., Keming, E., Bilong Bilong, C. F., Couloux, A., … Denys, C. (2012). Systematics and diversification of *Praomys* species (Rodentia: Muridae) endemic to the Cameroon Volcanic Line (West Central Africa). *Zoologica Scripta*, 41, 327–345.

Moritz, C., Patton, J. L., Schneider, C. J. & Smith, T. B. (2000). Diversification of rainforest faunas: An integrated molecular approach. *Annual Review of Ecology and Systematics*, 31, 533–563.

Nicolas, V., Verheyen, E., Verheyen, W., Hulselmans, J., Dillen, M., Akpatou, B., … Colyn, M. (2005). Systematics of African lowland rainforest *Praomys* (Rodentia, Muridae) based on molecular and craniometrical data. *Zoological Journal of the Linnean Society*, 145, 539–553.

Nicolas, V., Bryja, J., Akpatou, B., Konečný, A., Lecompte, E., Colyn, M., … Granjon, L. (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, 5118–5134.

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

Ogilvie, H. A., Bouckaert, R. R. & Drummond, A. J. (2017). StarBEAST2 brings faster species tree inference and accurate estimates of substitution rates. *Molecular Biology and Evolution*, 34, 2101–2114.

Phillips, S. J., Anderson, R. P. & Schapire, R. E. (2006). Maximum entropy modelling of species geographic distributions. *Ecological Modelling*, 190, 231–259.

Plana, V. (2004). Mechanisms and tempo of evolution in the African Guineo–Congolian rainforest. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 359, 1585–1594.

Plumptre, A. J., Davenport, T. R. B., Behangana, M., Kityo, R., Eilu, G., Ssegawa, P., … Moyer, D. (2007). The biodiversity of the Albertine Rift. *Biological Conservation*, 134, 178–194.

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

R Core Team (2016). *R: A language and environment for statistical computing.* R Foundation for Statistical Computing, Vienna, Austria. Available at: http://www.R-project.org/.

Robbrecht, E. (1996). Geography of African Rubiaceae with reference to glacial rain forest refuges. In L. J. G. van der Maesen, X. M. van der Burgt & J. M. van Medenbach de Rooy (Eds.), *The biodiversity of African plants* (pp. 564-581). Netherlands: Kluwer, Dordrecht.

Ronquist, F. & Huelsenbeck, J. P. (2003). MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics*, 19, 1572–1574.

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

Stebbins, G. L. (1974). Flowering plants: evolution above the species level. Cambridge, MA: Harvard University Press.

Sukumaran, J. & Knowles, L. L. (2017). Multispecies coalescent delimits structure, not species. *Proceedings of the National Academy of Sciences*, 114, 1607–1612.

Tamura, K., Stecher, G., Peterson, D., Filipski, A. & Kumar, S. (2013). MEGA6: molecular evolutionary genetics analysis version 6.0. *Molecular Biology and Evolution*, 30, 2725–2729.

Van der Straeten, E. & Kerbis Peterhans, J. C. (1999). *Praomys degraaffi*, a new species of Muridae (Mammalia) from central Africa. *South African Journal of Zoology*, 34, 80–90.

Yang, Z. & Rannala, B. (2014). Unguided species delimitation using DNA sequence data from multiple loci. *Molecular Biology and Evolution*, 31, 3125–3135.

Supporting information

Additional Supporting Information may be found in the online version of this article:

Appendix S1 List of specimens used in this study with the details of their localities, museum numbers, collectors, and available sequences.

Appendix S2 Additional figures and tables (Figs. S2.1, S2.2, S2.3, S2.4, Tables S2.1, S2.2).

Appendix S3 Detailed information about methods used for data analysis

Appendix S4 Complete results of BPP species delimitation

Data availability

Sequences were submitted to the GenBank database with accession codes MK422959-MK423164 for *CYTB*, MK511453-MK511555 and MK518347 for *ACPT*, MK511668-MK511781 and MK518346 for *CARHSP*, MK511556-MK511667 for *GAD2*. GenBank numbers and museum numbers of specimens are given in Appendix S1.

Biosketch

Daniela Mizerovská is a PhD student supervised by **Josef Bryja** at the Institute of Vertebrate Biology of the Czech Academy of Sciences, and this paper is based on her master thesis. **Violaine Nicolas** is researcher and curator of small mammals at MNHN in Paris and **Terrence Demos** is postdoctoral researcher at FMNH in Chicago. They all share interest in the evolutionary diversification of African small mammals. They use molecular, morphological, and distributional data to infer historical biogeography, phylogeography, and species limits.

Author contributions: VN, TD, EV and JB conceived the ideas; TD, JCK, JK, ADM, PK, AD, AL, EV, CD, MC, RS and JB collected samples in the field; DM, VN, TD, JK and JB genotyped the material; DM, VN and JB analysed the data; and JB, DM, VN and TD wrote the first version of the manuscript. All authors provided comments to the final version of the manuscript.

All figure legends should be standalone – mentioning the geographic sample and study species (where needed)

Figure legends

Fig. 1: Distribution of mitochondrial diversity based on 946 genotyped individuals of (a) *P. degraaffi* and *P. minor*; (b) *P. mutoni*; (c) *P. jacksoni* clade IV; (d) *P. jacksoni* sensu stricto, i.e. clades I and II. Colours and names of clades correspond to Fig. 2. Major biogeographic regions relevant to this study are schematically shown on panel (b). ARM = Albertine rift mountains, KH = Kenyan highlands.

Fig. 2: Mitochondrial phylogeny of the *P. jacksoni* complex. Bayesian tree based on 221 unique *CYTB* sequences is shown. Numbers above branches show posterior probability from MRBAYES/bootstrap support from RAxML for major nodes.

Fig. 3: Haplotype network of unique *CYTB* sequences of *P. jacksoni* sensu lato (i.e. including *P. minor* and *P. jacksoni* clade IV). The length of connecting lines correspond to the number of substitutions. Colours and names of taxa correspond to Fig. 2.

Fig. 4: Analysis of *CYTB* diversity by the 'GeneHubs' algorithm. The genetic hub locations for each species or haplogroup within species (shown by different colours) are indicated by asterisks. The color intensity indicates proximity to the hotspot of mtDNA variation, with the genetic hub being the most intense.

Fig. 5: Divergence dating of the species tree inferred using a multi-species coalescent approach in STARBEAST2. The numbers in circles are TMRCAs of particular clades. PP = posterior probability. (a) based on combined mitochondrial and nuclear dataset; (b) nuclear dataset only.

Fig. 6: Species distribution predicted in MAXENT for five species in the *P. jacksoni* complex. For each taxon, the large panel shows the environmental suitability in current conditions, while small panels show models for LIG (Last interglacial, ca. 120 000 – 140 000 before present) and LGM (Last glacial maximum, ca. 21 000 before present). Lighter colour indicates higher probability of suitable climatic conditions based on 19 BIOCLIM variables. Green dots represent genotyped records of particular taxa used for the construction of models.

Embedded figures

Fig. 1

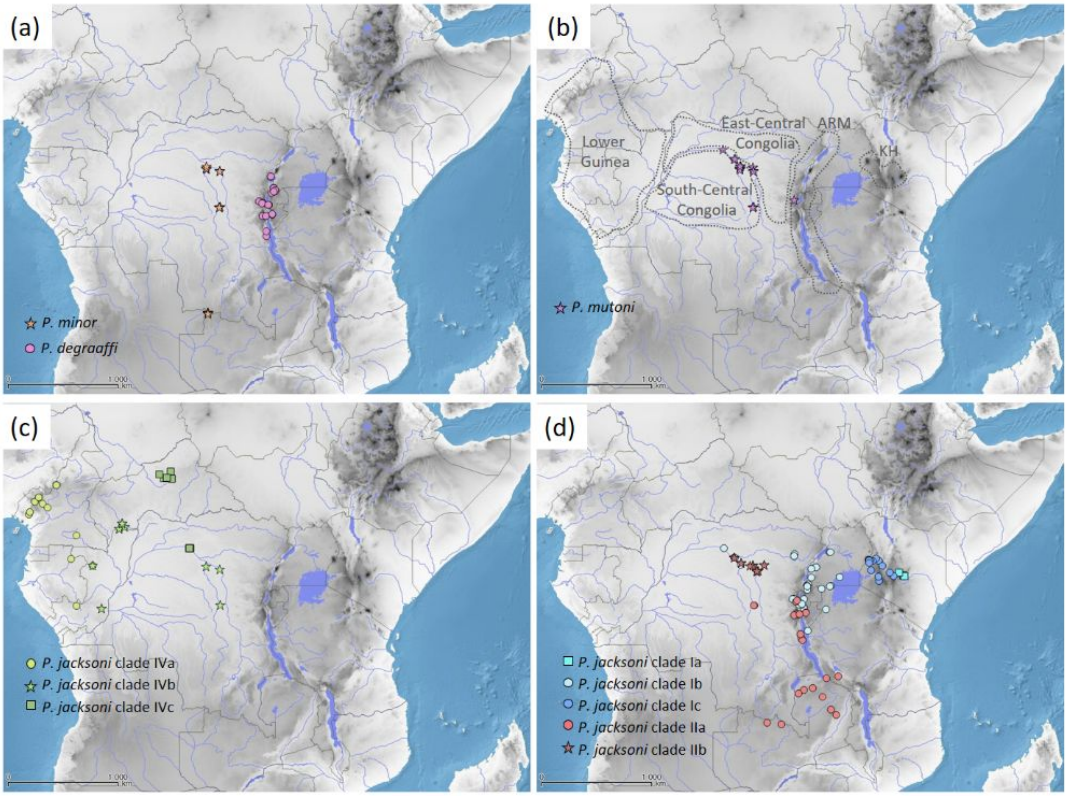
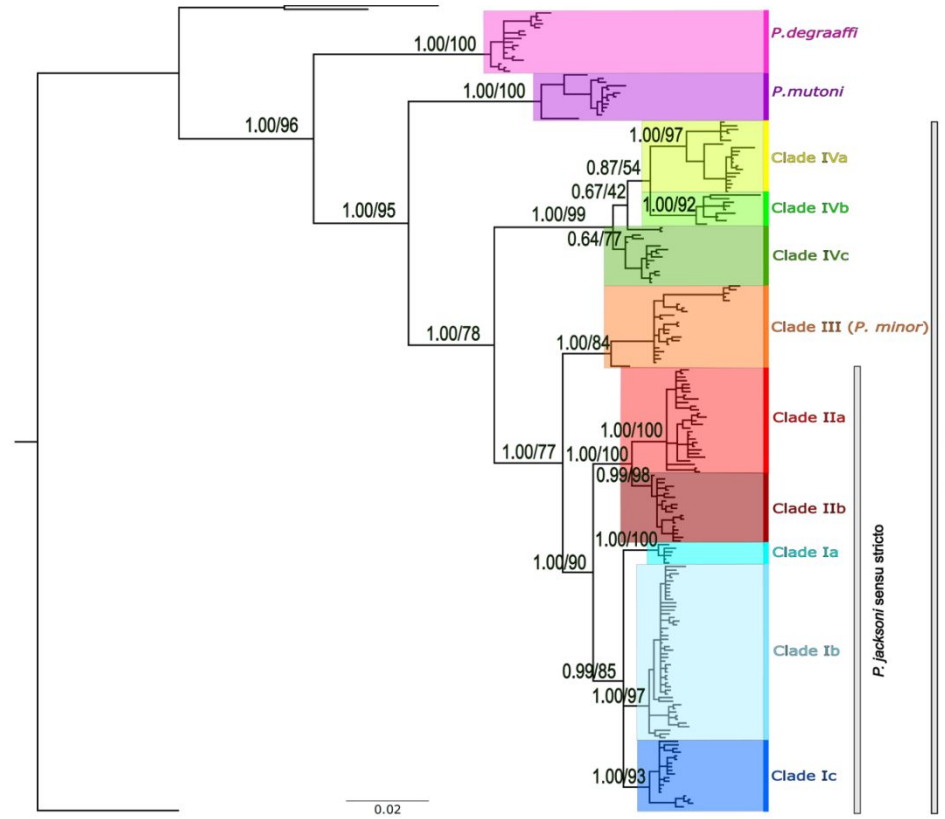
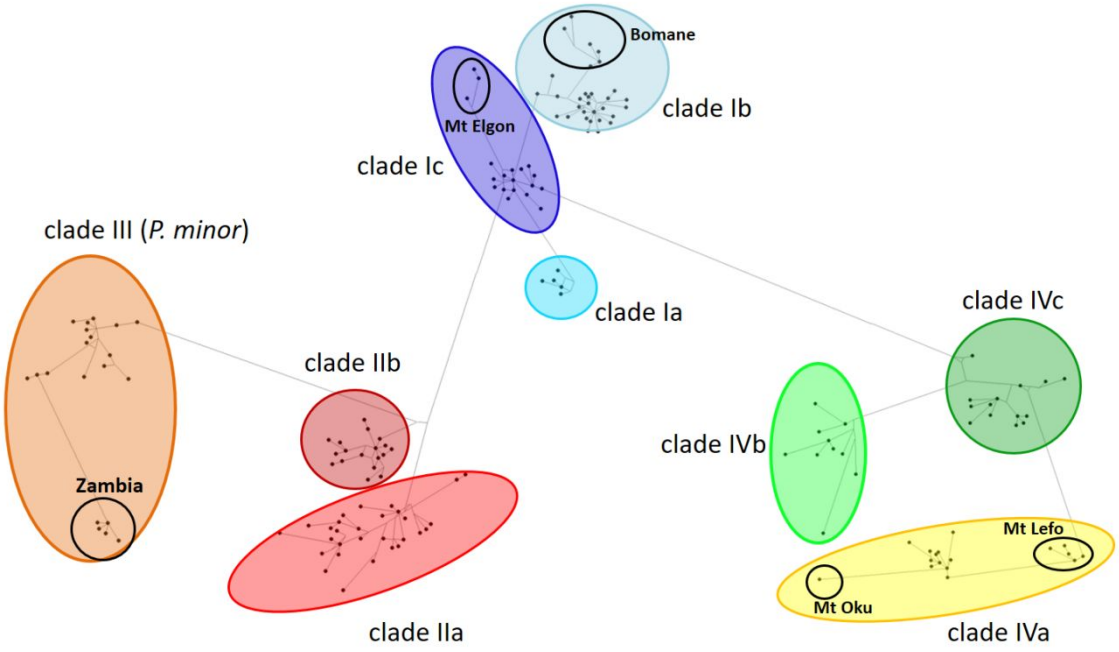


Fig. 2



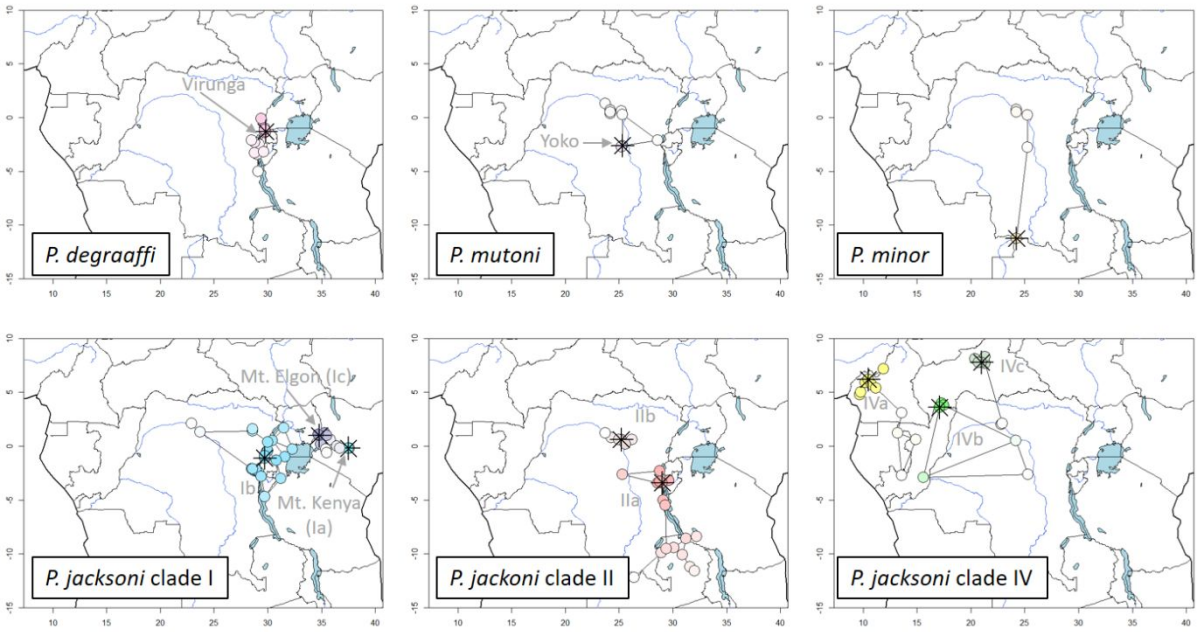
694

695 **Fig. 3**



696

697 **Fig. 4**



What figure is this?

Fig. 5

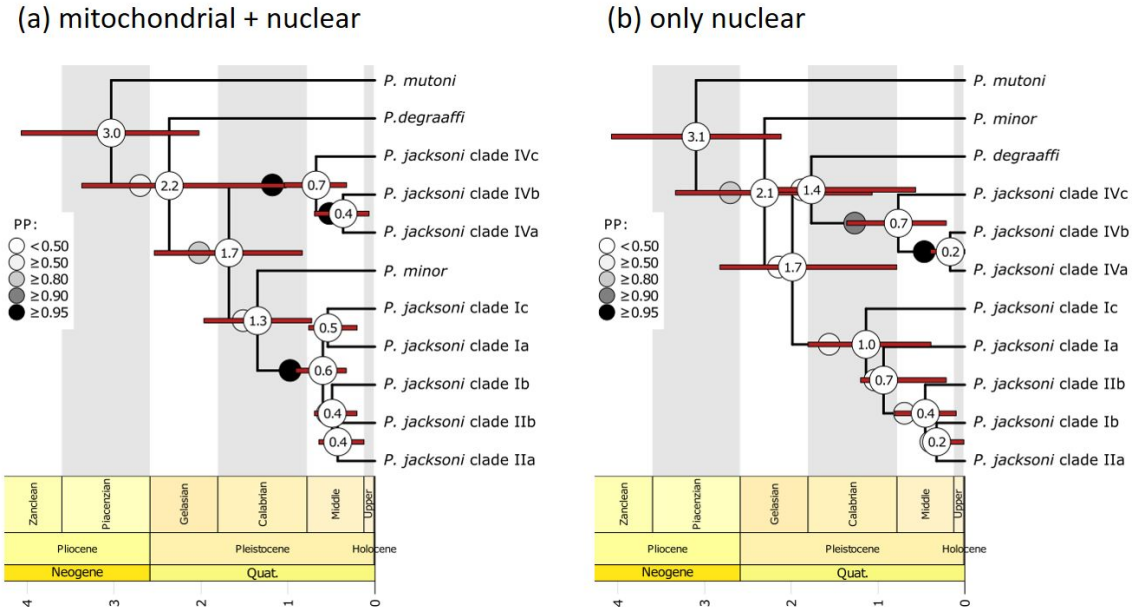
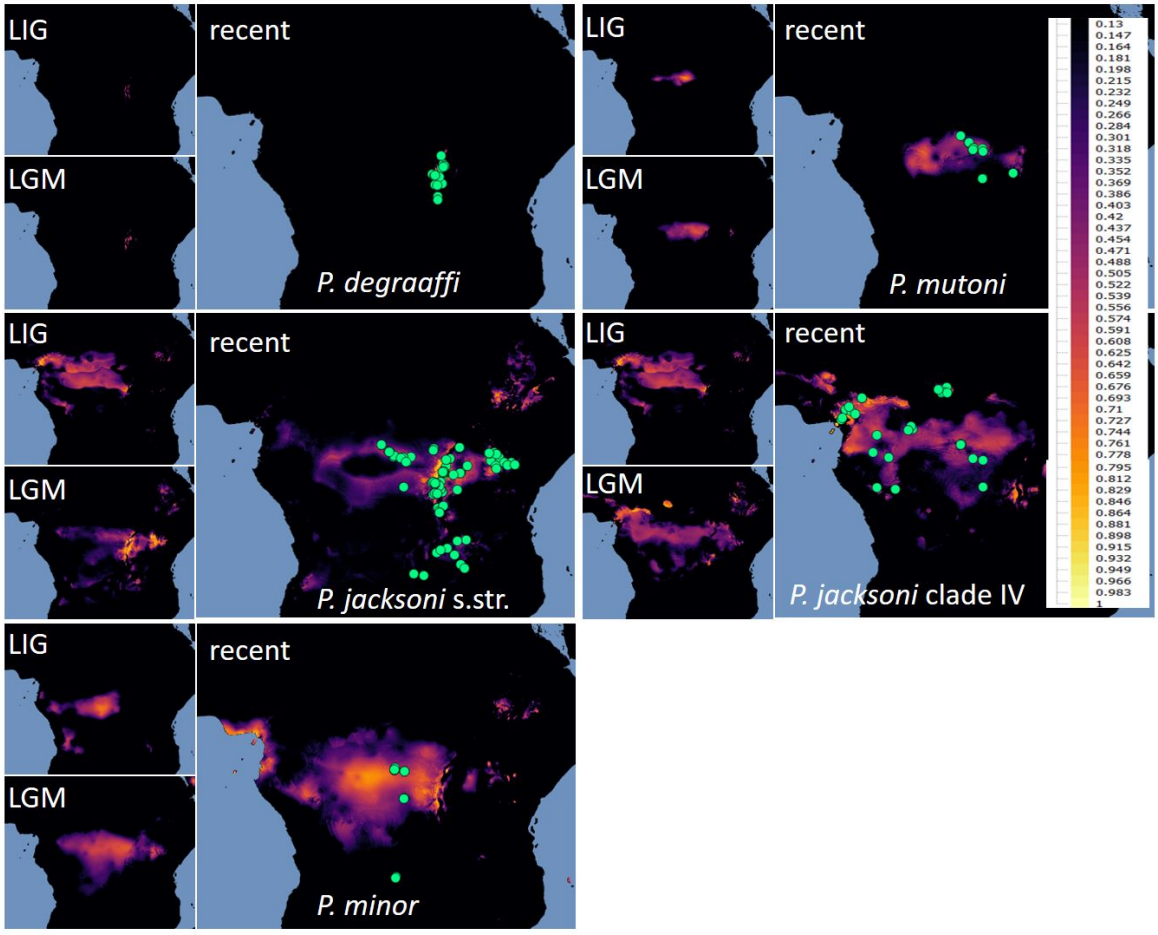

(a) mitochondrial + nuclear
(b) only nuclear
P. mutoni
P.degraaffi
P. jacksoni clade IVc
P. jacksoni clade IVb
P. jacksoni clade IVa
P. minor
P. jacksoni clade Ic
P. jacksoni clade Ia
P. jacksoni clade Ib
P. jacksoni clade IIb
P. jacksoni clade IIa
PP:
<0.50
≥0.50
≥0.80
≥0.90
≥0.95
Zanclean
Piacenzian
Gelasian
Calabrian
Middle
Upper
Pliocene
Pleistocene
Holocene
Neogene
Quat.

Fig. 6


LIG
LGM
recent
P. degraaffi
P. mutoni
P. jacksoni s.str.
P. jacksoni clade IV
P. minor