The phylogeny of the African wood mice (Muridae, Hylomyscus) based on complete mitochondrial genomes and five nuclear genes reveals their evolutionary history and undescribed diversity

Violaine Nicolas, Pierre-Henri Fabre, Josef Bryja, Christiane Denys, Erik Verheyen, Alain Didier Missoup, Ayodeji Olayemi, Pionus Katuala, Akaibe Dudu, Marc Colyn, et al.

The phylogeny of the African wood mice (Muridae, *Hylomyscus*) based on complete mitochondrial genomes and five nuclear genes reveals their evolutionary history and undescribed diversity

Nicolas Violaine[1], Fabre Pierre-Henri[2], Bryja Josef[3], Denys Christiane[1], Verheyen Erik [4], Missoup Alain-Didier[5], Olayemi Ayodeji[6], Katuala Pionus[7], Dudu Akaibe[7], Colyn Marc[8], Kerbis Peterhans Julian[9,10], Demos Terrence[9]

(1) Institut de Systématique, Evolution, Biodiversité (ISYEB), Muséum national d'Histoire naturelle, CNRS, Sorbonne Université, EPHE, Université des Antilles, CP51, 75005 Paris, France;
(2) Institut des Sciences de l'Evolution (ISEM, UMR 5554 CNRS-IRD-UM), Université de Montpellier, Place E. Bataillon CC 064 - 34095, Montpellier Cedex 5, France ;
(3) Institute of Vertebrate Biology of the Czech Academy of Sciences, Květná 8, 603 65, Brno, Czech Republic and Department of Botany and Zoology, Faculty of Science, Masaryk University, Kotlářská 2, 611 37 Brno, Czech Republic;
(4) Royal Belgian Institute for Natural Sciences, Operational Direction Taxonomy and Phylogeny, 1000, Brussels, Belgium;
(5) Zoology Unit, Laboratory of Biology and Physiology of Animal Organisms, Faculty of Science, University of Douala, Douala, Cameroon;
(6) Natural History Museum, Obafemi Awolowo University, HO 220005 Ile Ife, Nigeria
(7) Animal Ecology and Resource Management, Laboratory (LEGERA), University of Kisangani, B.P. 2012, Kisangani, Democratic Republic of Congo;
(8) Université de Rennes 1, CNRS, UMR 6553 Ecobio, Station Biologique, 35380 Paimpont, France;
(9) Science & Education, Field Museum of Natural History, 60605, Chicago, IL, USA
(10) College of Arts & Sciences, Roosevelt University, 430 S Michigan, Chicago, IL 60605, USA

Corresponding author:
Nicolas Violaine: violaine.colin@mnhn.fr
Institut de Systématique, Evolution, Biodiversité (ISYEB), Muséum national d'Histoire naturelle, CNRS, Sorbonne Université, EPHE , Université des Antilles, 47 rue Cuvier,

CP51, 75005 Paris, France

Abstract

Wood mice of the genus *Hylomyscus*, are small-sized rodents widely distributed in lowland and montane rainforests in tropical Africa, where they can be locally abundant. Recent morphological and molecular studies have increased the number of recognized species from 8 to 18 during the last 15 years. We used complete mitochondrial genomes and five nuclear genes to infer the number of candidate species within this genus and depict its evolutionary history. In terms of gene sampling and geographical and taxonomic coverage, this is the most comprehensive review of the genus *Hylomyscus* to date. The six species groups (*aeta*, *alleni*, *anselli*, *baeri*, *denniae* and *parvus*) defined on morphological grounds are monophyletic. Species delimitation analyses highlight undescribed diversity within this genus: perhaps up to 10 taxa need description or elevation from synonymy, pending review of type specimens. Our divergence dating and biogeographical analyses show that diversification of the genus occurred after the end of the Miocene and is closely linked to the history of the African forest. The formation of the Rift Valley combined with the declining global temperatures during the Late Miocene caused the fragmentation of the forests and explains the first split between the *denniae* group and remaining lineages. Subsequently, periods of increased climatic instability during Plio-Pleistocene probably resulted in elevated diversification in both lowland and montane forest taxa.

Keywords

Biogeography ; Mammals ; Rodents ; Speciation ; Taxonomy ; Tropical Africa.

1- Introduction

Wood mice of the genus *Hylomyscus* Thomas, 1926, are small-sized rodents belonging to the family Muridae (Wilson et al., 2017). They are distributed in lowland and montane rainforests in tropical Africa, where they can be locally abundant (Happold, 2013; Nicolas and Colyn, 2003; Rosevear, 1969). Morphological and molecular data support the monophyly of the genus (Lecompte et al., 2008; Lecompte et al., 2002a; Lecompte et al., 2002b; Nicolas et al., 2006; Nicolas et al., 2012) and its placement within the Praomyini tribe, with the genus *Heimyscus* as its nearest relative (Aghova et al., 2018; Lecompte et al., 2008; Schenk et al., 2013; Steppan and Schenk, 2017). Based on external and craniodental morphology,

Hylomyscus species were separated into six species groups: *aeta*, *alleni*, *anselli*, *baeri*, *denniae* and *parvus* (Carleton et al., 2006). Within these groups, individual species are morphologically similar and, as a result, the taxonomic status of many populations remains a subject of debate (Carleton et al., 2015; Robbins et al., 1980; Rosevear, 1969; Rosevear, 73 1966). From just eight species recognized by Musser & Carleton (2005), recent morphological, morphometric and molecular studies have increased the number of recognized species to 18 (Carleton et al., 2015; Carleton et al., 2006; Carleton and Stanley, 2005; Nicolas et al., 2010; Nicolas et al., 2006; Nicolas et al., 2008b; Wilson et al., 2017), and more species surely await description (Nicolas et al., 2006; Nicolas et al., 2012; Olteanu et al., 2013). The *baeri* group is composed of a single West African lowland species, *H. baeri*. The *aeta* group contains two species (Wilson et al., 2017): *H. aeta* (found both in lowland and montane forests of Central Africa) and *H. grandis* (endemic to montane forest of Mt Oku in Cameroon). The taxonomy of *H. aeta* is still unclear and it may represent a species complex: strong phylogeographic structure was observed within this species and at least one cryptic species has been reported, but not yet described within it (Kennis, 2012; Nicolas et al., 2006; Nicolas et al., 2012; Olteanu et al., 2013). The *parvus* group is currently considered to contain only one lowland species from Central Africa (Wilson et al., 2017), but molecular studies suggest the presence of several cryptic species within it, not yet formally described (Nicolas et al., 2006; Nicolas et al., 2012; Olteanu et al., 2013). The *alleni* group is widely distributed in West and Central Africa lowland forest and includes six nominal species (Monadjem et al., 2015; Wilson et al., 2017): *H. alleni*, *H. stella*, *H. walterverheyeni*, *H. simus*, *H. pamfi, H. carillus* and *H. kaimosae*. In most revisions of the genus *H. kaimosae* is considered as a synonym of *H. stella*, but in the last revision Wilson et al. (2017) considered it as a species based on the molecular data of Nicolas et al. (2006; 2012). However, the specimens named *kaimosae* in the study of Nicolas et al. (2006; 2012) were misidentified and corresponds to *H. arcimontensis*. The *alleni* group has a complicated taxonomic history because the holotype of *alleni* is a juvenile specimen with immature dentition and imprecise geographical origin, which compromises the utility of the species description and has lead several authors to consider this species as essentially unidentifiable (Rosevear, 1969). At least two cryptic species have been reported, but not yet described, within the *alleni* group, and *H. simus* probably represents a complex of species (Nicolas et al., 2006; Nicolas et al., 2012; Olteanu et al., 2013). The *denniae* group is composed of three East African allopatric montane forest specialists: *H. denniae* (Ruwenzori Mts), *H. endorobae* (Kenyan Highlands) and *H. vulcanorum* (central and southern Albertine Rift). Lastly, the *anselli* group is composed of

four allopatric montane forest species: *H. anselli* (Zambia), *H. arcimontensis* (Tanzania, Malawi), *H. heinrichorum* (Angola) and *H. kerbispeterhansi* (Kenya). Four new species are being described within this group (Kerbis Peterhans et al., in review).

No phylogenetic study of the genus including all currently recognized species is available. Such a study would allow us to test the robustness of each species group delimited by morphology, and to better assess the total number of *Hylomyscus* species. This is important not only for general knowledge of systematics, taxonomy and the distribution of biodiversity, but also for ecological and virological studies. *Hylomyscus* is often among the most abundant members of the rodent communities in tropical African forest (Denys et al., 2009; Duplantier, 1989; Kennis, 2012; Nicolas and Colyn, 2003; Olayemi and Akinpelu, 2008) and it is known to carry several hantaviruses (Klempa et al., 2006; Sudi et al., 2018; Tesikova et al., 2017) and arenaviruses (Coulibaly-N'Golo et al., 2011; Olayemi et al., 2016), i.e. virus groups potentially harmful also to humans. Because it contains both lowland and montane forest specialists, the genus is widely distributed in tropical Africa, and an improved knowledge of its evolutionary history is key for our understanding of the dynamics of African rainforests.

In this study we analyze the most comprehensive genetic dataset of the genus *Hylomyscus* to date including both complete mitogenomes and five nuclear genes. This allows us to 1) clarify its systematics and taxonomy, and 2) investigate the evolutionary history of the genus, particularly the factors shaping its genetic diversity and geographical distribution.

2- Material and methods

2-1 Sampling

We obtained genetic data from 141 *Hylomyscus* specimens (Supplementary Data S1), representing all six species groups (*baeri*, *aeta*, *parvus*, *alleni*, *denniae* and *anselli*). Our sampling includes all but two currently recognized and described species of the genus. We could not obtain DNA of *H. carillus* and *H. grandis*, only known from Angola, and from montane forest of Mt. Oku in Cameroon, respectively. We included one virtual topotype of *H. stella kaimosae* from Kakamega Forest and several taxa representing candidate species identified by previous molecular studies (Nicolas et al., 2006; Nicolas et al., 2012; Olteanu et al., 2013). Candidate species are populations, for which there is insufficient evidence of species status, and that have not yet received a formal name (Padial and De la Riva, 2010). We included in our analysis the candidate species *H. sp1, H. sp2* and *H. sp7* of Nicolas et al. (2012), *H.* cf. *anselli* (Kerbis Peterhans et al., in review), the three clades of *H. parvus* (here called *H. parvus1*, *H. parvus2*, *H. parvus3*), and the two clades of *H. simus* (here

called *H. simus1* and *H. simus2*) identified in Nicolas et al. (2006) and/or Nicolas et al. (2012). For each species, two to 19 specimens were sequenced. Whenever possible, conspecific individuals were chosen from geographically distant localities in order to take into account intra-specific genetic variability.

We also obtained genetic data for 21 additional specimens belonging to other rodent genera used as outgroups in our phylogenetic and divergence time analyses (Supplementary Data S1).

Specimens were collected during recent fieldwork performed by the authors and their collaborators. All fieldwork complied with environmental regulations in the respective African countries and sampling was carried out in accordance with local legislation. Detailed information on specimens is presented in Supplementary Data S1.

2-2 DNA extraction, amplification and sequencing

Total genomic DNA was extracted from ethanol-preserved muscle, liver or heart by either the Cetyl Trimethyl Ammonium Bromide method (Winnepenninckx et al., 1993) or the NucleoSpin 96 Tissue Kit (Macherey-Nagel, Germany) and the automatic pipetting robot Eppendorf epMotion 5075 TMX following the manufacturer's manual.

The complete mitochondrial genome was obtained by overlapping three DNA fragments amplified by long-range PCRs and the HotStart LongAmp Taq DNA polymerase (New England BioLabs Inc., Ipswich MA). The primers used for these PCRs are presented in Supplementary Data S2. The amplicons were sequenced using the Ion Torrent Personal Genome Machine (PGM) sequencing platform with two level multiplexing (Hinsinger et al., 2016) at the "Service de Systématique Moléculaire" of the MNHN. The reads obtained from each long-range PCR were trimmed by an error probability algorithm and assembled using the mapping to reference algorithm in Geneious 9.0.4 version (Kearse et al., 2012). The reference used for mapping was the mitochondrial genome of *Mastomys natalensis* (NC036995).

For 11 samples, libraries were prepared following the protocol described in Tilak et al. (2015). These libraries were pooled and sequenced without enrichment as single end reads on Illumina HiSeq 2000 lanes at the GATC-Biotech company (Konstanz, Germany). Raw 101-nt reads were imported in Geneious (Kearse et al., 2012) and adaptor fragments were removed by the "trim ends" utility. Then, a mapping of the reads on the phylogenetically closest available mitochondrial genome was performed for each species. The following mapping parameters were used in the Geneious read mapper: a minimum of 24 consecutive nucleotides (nt) perfectly matching the reference, a maximum 5% of single nt mismatch over the read

length, a minimum of 95%-nt similarity in overlap region, and a maximum of 3% of gaps with a maximum gap size of 3-nt. Iterative mapping cycles were performed in order to elongate the sequence when the complete mitogenome was not recovered after the initial mapping round. A high-quality consensus was generated and the circularity of the mitogenome was verified by the exact superimposition of the 100-nt at the assembly extremities.

Five unlinked autosomal nuclear introns were also amplified: GAD2-1 and JMJD5-2 using primers from Igea et al. (2010); and ABHD11-5, ACOX2-3 and ACPT-4 using primers from Salicini et al. (2011). These nuclear introns were selected because they were previously used by Demos et al. (2014a; 2014b) and showed to successfully delimit several closely related East African *Hylomyscus* species. The amplicons were sequenced using the Ion Torrent Personal Genome Machine (PGM) sequencing platform with two level multiplexing (Hinsinger et al., 2016). The reads were trimmed by error probability algorithm and assembled using the mapping to reference algorithm in Geneious (Kearse et al., 2012). The reference used for mapping were KF810445 for GAD2-1, KF810493 for JMJD5-2, KF810266 for ABHD11-5, KF810328 for ACOX2-3 and KF810387 for ACPT-4.

All new sequences were deposited into GenBank (Supplementary Data S1).

As far as possible nuclear alleles were resolved based on reads obtained through the PGM sequencing. When it was not possible to deduce alleles from next generation sequencing (no overlaps between reads with heterozygotes positions) and for sequences downloaded from GenBank nuclear alleles were resolved statistically using PHASE 2.1 (Stephens et al., 2001). Input files for PHASE were assembled using the SEQPHASE web server (Flot, 2010).

2-3 Phylogenetic analyses

We first reconstructed a mitochondrial phylogeny. Using Geneious, we aligned newly assembled *Hylomyscus* mitochondrial genome sequence with the sequences of three other members of the Praomyini tribe (*Heimyscus fumosus*, *Mastomys natalensis* and *Praomys rostratus*) used as outgroups. Sequence data from the 13 protein coding genes were translated into amino acids and inspected for deletions, insertions and premature stop codons to prevent inclusion of paralogous sequences. Alignments for all data sets were inspected visually and determined to be unambiguous. We used PartitionFinder v2.1.1 (Lanfear et al., 2017) to find the best partition schemes and nucleotide substitution models, with the greedy algorithm and considering 43 potential partitions: all tRNAs (each individual tRNA alignment was too short to be informative), 12S rRNA, 16S rRNA, the control region, and the 13 protein-coding genes

each subdivided into first, second, and third codon position. We used linked branch lengths and the Bayesian information criterion (BIC) to select the best-fitting model and partition scheme, which were then used for tree reconstructions. The best supported scheme inferred 12 partitions (Supplementary Data S3). Bayesian tree reconstruction was performed with MrBayes v3.2.6 (Ronquist et al., 2012) using the best-fit partition scheme and nucleotide substitution models determined by PartionFinder. Two independent runs of four Metropolis-coupled Markov chains (MCMC) for Monte Carlo simulations were run for 5,000,000 generations, with parameters and trees sampled every 100 generations. Convergence was checked using Tracer 1.5 (Rambaut et al., 2013). For each run, the first 25% of sampled trees were discarded as burn-in. Bayesian posterior probabilities (PP) were used to assess node support of the Markov chain Monte Carlo (MCMC) tree. Maximum likelihood analyses were performed with RAxML 8.0 (Stamatakis, 2014). RAxML only implements GTR-based models of nucleotide substitution; thus we used the GTRGAMMA substitution model and the partition scheme determined by PartitionFinder. Robustness of the nodes were assessed using the rapid bootstrap procedure with 1000 replications.

We also constructed a nuclear phylogeny using the five concatenated nuclear introns. Using PartitionFinder 2.1.1 and BIC, the best scheme supported two partitions (Supplementary Data S3). Phylogenetic analyses were performed in RAxML and MrBayes as described above.

2-4 DNA divergence within species

The average number of substitution per site (K; p-distance) within each species was calculated using MEGA X (Kumar et al., 2018).

2-5 Species delimitation

2-5-1 ABGD on mtDNA dataset

The Automatic Barcode Gap Discovery (ABGD) algorithm proposed by Puillandre et al. (2011) was used to identify gaps in the distribution of pairwise nucleotide distances. ABGD categorized the mtDNA sequences into MOTUs based on the barcode gap. The ABGD algorithm detects the largest significant gap in the distribution of pairwise nucleotide distances and uses it for partitioning the dataset. Gap detection is then recursively applied to previously obtained groups to get finer partitions until there is no further partitioning.

The ABGD analysis was conducted on the web interface (www.abi.snv.jussieu.fr/public/abgd/), with both the Jukes-Cantor (JC69) and *p*-distance models and a relative gap width of X = 1. Here we did not use the K2P model because

assignment of this model for barcoding is questionable and it is not appropriate for closely related sequences (Srivathsan and Meier, 2012). The prior for the maximum value of intraspecific divergence (*P*) was set to a range of 0.001 to 0.1 and the number of recursive steps was set to 10.

2-5-2 Bayesian coalescent species-tree approaches

We carried out delimitation analyses using the most recent version of Bayesian Phylogenetics and Phylogeography (BPP) v.3.4 (Yang, 2015). This method employs a multispecies coalescent model to evaluate support for alternative hypotheses of species delimitation and species tree phylogeny in a Bayesian framework, while accounting for incomplete lineage sorting and gene tree/species tree conflict. Three species delimitation scenarios were tested: 1) a scenario with 24 candidate species: 24 MOTUs (Molecular Taxonomic Units) identified in the ABGD analysis with the recursive partition and JC69 distance; 2) a scenario with 40 candidate species: maximum number of MOTUs recovered in ABGD analyses; 3) a scenario with 25 candidate species (see the results section for an explanation of this choice). We tested the distinctiveness of their gene pools using the nDNA dataset, and all nuclear genes were phased.

Bayesian species delimitation was conducted using the joint species delimitation and species-tree estimation model (Yang, 2015). 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 distinct taxa (species). We consider speciation probability values >0.95 as strong support for a putative speciation event. BPP 3.4 uses inverse-gamma priors for the population size parameters (θ) and the age of the root in the species tree (τ_0). Based on the variability obtained in our dataset we assigned the inverse-gamma priors $\theta \sim$ IG(3, 0.001) and $\tau_0 \sim$ IG(3, 0.03). We allowed the loci to have different rates (locus rate = 1 and Dirichlet distribution). Each analysis was run using both algorithm 0 and algorithm 1 (Yang and Rannala, 2010). We ran the rjMCMC analyses for 35×10^4 generations with a burn-in period of 7×10^3.

2-6 Divergence time estimates and biogeographic analyses

Divergence times were estimated from: 1) the combined partial mitochondrial (parts of the 12S, 16S and control region for which the alignment was doubtful were removed) + nuclear supermatrix to provide a temporal framework for *Hylomyscus* evolution, as well as on 2) the nuclear matrix only, and 3) the mitochondrial one (without poorly aligned rRNA and control

region). In all molecular dating analyses we fix the topology using our MrBayes consensus tree. We retained one specimen per candidate species defined by ABGD and BPP analyses. Molecular dating analyses were performed with PhyloBayes v4.1 (Lartillot et al., 2009). We employed PhyloBayes to estimate the divergence dates on 1) the supermatrix, 2) the mitochondrial matrix, and 3) the nuclear matrix.

PhyloBayes implements the CAT+GTR mixture model that is suited to study our heterogeneous mito-nuclear supermatrix. This mixture model allows to identify class of nucleotidic sites with heterogeneous substitution rates including potential case of heterotachy (Philippe et al. 2005a, b). The CAT+GTR mixture model was used to precisely estimate divergence dates while accounting for changes in evolutionary rate over time and allowing for mixture models of sequence evolution to account for heterogeneities in patterns of evolution. PhyloBayes use a Bayesian relaxed molecular clock with rate autocorrelation along the branches of the tree (Thorne et al., 1998; Thorne and Kishino, 2002; Lepage et al. 2007). We assumed the Birth-Death model of speciation as recommended by Ritchie et al. (2017) as tree priors. For each dataset (supermatrix, nuclear and mitochondrial) two Markov chain Monte Carlo analyses were run with PhyloBayes for 10,000 cycles (ca., 8,000,000 generations) with trees sampled every five cycles after discarding the first 1,000 as burn-in. Convergence was ensured when the maximum difference in bipartition frequencies as estimated by the two chains was <0.1. Calibrations were set as uniform priors (see their description below).

To calibrate the phylogeny we selected 3 fossil uniform constraints as described from previous studies described in Aghova et al. (2018), see Table 2 in their manuscript for the used priors and a thorough discussion of the Murinae fossil record. In order to take into account uncertainties in the phylogenetic position of these fossils, all constraints were set using hard minimum bounds and soft upper bounds using a lognormal prior, as suggested by recent paleontological studies (Benton and Donoghue, 2007; Benton *et al.*, 2009; Parham *et al*., 2012). We used the following constraints: 1) The most recent common ancestor (MRCA) among the core Murinae was used to calibrate the *Mus* / *Rattus* divergence. The fossil evidence (Jacobs and Flynn, 2005) suggests the acquisition of a major synapomorphy (full fusion of the maxillary lingual cusps with the medial and labial cusps) within the core murine taxa approximately 13.8 Mya (Kimura et al. (2015). This minimum boundary corresponds to the Middle Miocene split between *Antemus chinjiensis* and *Progonomys*. Uniform priors in PhyloBayes were set as follow: soft upper bound = 25.29 Mya; hard minimum bound = 12.5 Mya. 2) We used the divergence of Arvicanthini and Murini lineages by putting as prior the age of their most recent common ancestor from the late Miocene, i.e. cf. *Karnimata* sp. (Aghova et al., 2018; Kimura

et al., 2015; Kimura et al., 2013). Uniform priors in PhyloBayes were set as follow : soft upper bound = 25.42 Mya ; hard minimum bound = 8.52 Mya. 3) We used the most recent common ancestors of Millardini, Otomyini and Arvicanthini from the late Miocene using *Karnimata darwini* (Aghova et al., 2018; Kimura et al., 2015; Kimura et al., 2013). Uniform priors in PhyloBayes were set as follow: soft upper bound = 25.37 Mya; hard minimum bound = 10.47 Mya.

Biogeographical reconstructions were performed using the chronogram generated by PhyloBayes on the mito-nuclear supermatrix and the use of the Dispersal–Vicariance Analysis-like model (DIVA-like, likelihood interpretation of DIVA) and the Dispersal-Extinction Cladogenesis model (DEC) in RASP 4.02 (Yu et al., 2015). Several other models for biogeographic inference are available and it is now a common practice to choose between these models of range evolution using statistical model selection in BioGeoBEARS (Matzke, 2013). However, Ree & Sanmartin (2018) recently showed that there are conceptual and statistical problems in using statistical model selection to choose between models. The "J" parameter (i.e., jump dispersal) was not included in our analysis because it is not necessary for testing our hypothesis, and because concerns have been raised that this parameter can artificially inflate likelihood values resulting in a tendency towards explaining the data disproportionately by "jump dispersal" events (Ree & Sanmartín, 2018). As terminal biogeographical units, we used four main forest types defined by Fayolle et al. (2014): A- wet and moist forests of West Africa, B- dry forest of West Africa, C- wet and moist forests of Central Africa, and D- upland forest of East Africa, Zambia and Angola. We followed Table 1 to define in which biogeographical unit a given lineage is present. DIVA is known to be sensitive towards the exclusion of outgroup taxa (Kodandaramaiah, 2010). To overcome this bias we retained the sister group to the genus *Hylomyscus*, i.e. *Heimyscus fumosus* in the analyses. The maximum number of areas was set to two. For the DEC analysis the dispersal probability between areas was set to 1.0 for adjacent areas and to 0.7 for well-separated areas (i.e. A-D and A-C). The probability of ancestral areas was plotted in the form of pie-charts along the divergence time analysis tree.

3- Results

3-1 mtDNA phylogeny

Our mtDNA phylogenetic tree (Fig 1, Supplementary Data S4) is well resolved with all nodes above species level (except five) with maximal support (BP = 100, PP = 1.0). This tree

supports the monophyly of the genus *Hylomyscus* and of the six species groups previously defined on the basis of external and craniodental morphology by Carleton et al. (2006). The *anselli* and *parvus* species groups are sisters, and they form a monophyletic group with the *alleni* species group. The *aeta* and *baeri* species groups are also sisters. Finally, the *denniae* species group is sister to all other *Hylomyscus* clades.

The taxa *H. stella*, *H. walterverheyeni*, *H. simus1*, *H. simus2*, *H. pamfi*, *H.* cf. *alleni* as well as the candidate species *H. sp2* are well supported and cluster within the *alleni* species group. *Hylomyscus stella* and *H. walterverheyeni* are sister species. The two candidate species *H. simus1* and *H. simus2* are well supported sister groups. Moreover, they form a clade with *H. pamfi*. *Hylomyscus* cf *alleni* and *H. sp2* are sister species. High genetic variability is observed within *H. walterverheyeni* (K = 0.0248; Supplementary Data S5) and *H. pamfi* (K = 0.0143) with several well supported groups.

The species *H. kerbispeterhansi*, *H. arcimontensis*, *H. heinrichorum*, *H. anselli* and the candidate species *H. sp7* and *H.* cf. *anselli* cluster within the *anselli* species group. The nearly topotypic specimens of *H. anselli* from Zambia are the sister clade of *H. heinrichorum* from Angola. The *Hylomyscus* population from Mbizi Mts, south-western Tanzania, previously assigned to *H. anselli* (Carleton and Stanley, 2005), represents a distinct undescribed species here called *H.* cf. *anselli*. *Hylomyscus kerbispeterhansi* and *H.* cf. *anselli* are sister species, and they form a moderately supported clade with *H. sp7* (BP = 73; PP = 0.99). *Hylomyscus arcimontensis* is the sister clade of these three species. High genetic variability is observed within *H. sp7* (K = 0.0152) and *H. arcimontensis* (K = 0.017).

The three candidate species *H. parvus1*, *H. parvus2*, and *H. parvus3* are well supported and they significantly cluster together (BP = 99, PP = 1.0), *H. parvus2* and *H. parvus3* being sister species (BP = 92, PP = 1.0). High genetic variability is observed within *H. parvus1* (K = 0.0137) and *H. parvus2* (K = 0.0179).

The *aeta* species group is composed of *H. aeta*, within which there is high genetic variability (K = 0.0155), and the genetically distinct candidate species *H. sp1*.

The *denniae* species group comprises three species: *H. vulcanorum*, *H. denniae* and *H. endorobae*, the two first species being sister (but the bootstrap support of this node is relatively low: BP = 91).

3-2 nDNA phylogeny

The resolution of the concatenated nDNA phylogenetic tree (Fig. 2, Supplementary Data S6) is generally lower than that of mitogenomic tree. Five of the six species groups are recovered

in the nDNA tree (*alleni*, *anselli*, *aeta*, *baeri* and *denniae*) and the phylogenetic relationships between them are congruent with those observed with the mtDNA dataset. The species group *parvus* is not recovered in the nDNA tree, but support for this species group was already lower in the mtDNA tree. The two candidate species *H. parvus1* and *H. parvus2* cluster with high support, while the position of *H. parvus 3* is unresolved.

Within the *alleni* species group only the species *H. stella*, *H. simus* (including both *H. simus1* and *H. simus2*) and *H. pamfi* are recovered as monophyletic. Specimens of *H. walterverheyeni* cluster with *H. stella*. Phylogenetic resolution between species is low, with few nodes being well supported.

Within the *anselli* species group all recognized species and candidate species are recovered as monophyletic, and phylogenetic relationships between species are congruent with those observed at mtDNA tree, except for *H. arcimontensis* which is the sister clade of the group *H. kerbispeterhansi* + *H.* cf *anselli* in the nDNA dataset, but the support of this clade is moderate in the nDNA tree (BP = 90, PP = 1.0; compare Figs. 1 and 2).

The *aeta* species group is composed of *H. aeta* and *H. sp1,* and high genetic variability is observed within *H. sp1* with two highly supported clades.

The *denniae* species group comprises three species: *H. vulcanorum*, *H. denniae* and *H. endorobae*. Phylogenetic resolution between species is low, few nodes being well supported.

3-3 ABGD on the mtDNA dataset

ABGD categorized the mtDNA sequences into MOTUs based on the barcode gap. Initial partition with the two distance methods (JC69 and *p*-distance) produces 20 MOTUs with a prior intraspecific divergence value varying from 0.0010 to 0.0215 (Supplementary Data S7). This method detects one MOTU for each morphospecies or candidate species proposed by previous studies, except in the case of (1) *H. anselli* which is grouped with *H. heinrichorum*, and (2) *H.* cf. *anselli* which is grouped with *H. kerbispeterhansi* (Fig. 1).

In the recursive partitions the number of MOTUs varied from 23 to 39, and from 24 to 40 depending on the intraspecific divergence prior, for the *p*-distance and JC distance, respectively (Fig. 1). In the 23 MOTUs partition all morphospecies and candidate species proposed by previous studies are considered as distinct MOTUs, and *H. sp7* is split into two MOTUs (one MOTU with the specimens from Djabir, and one MOTU with the specimens from Yoko and Babusoko-Amadiabe). The 24 MOTUs partition is identical to the 23 MOTUs partition except that *H. pamfi* is split into two MOTUs (specimens from Benin, and specimens from Nigeria, respectively). When the number of MOTUs increases, the taxa *H. aeta*, *H.*

walterverheyeni, *H. arcimontensis, H. stella* and *H. kerbispeterhansi* are successively split into several MOTUs.

3-4 BPP on the nDNA dataset

When considering 24 candidate species in BPP analyses (24 MOTUs identified in the ABGD analysis with the recursive partition and JC69 distance), all analyses support the existence of 23 species (all specimens from *H. pamfi* are grouped in a single candidate species) with high posterior probability, and the posterior probability of each delimited species is also high (PP > 0.95; Supplementary Data S8; Fig. 2).

When considering 40 candidate species (maximum number of MOTUs recovered in ABGD analyses), BPP recognizes the 30 or 31-species delimitation scenario as the most probable one, but the posterior probabilities of these models are low (< 0.38; Supplementary Data S8, Fig. 2), and several lineages have posterior probabilities below 0.95 rejecting their recognition as distinct species. *Hylomyscus aeta* is considered as two distinct species with high posterior probabilities (Bioko - Mt Cameroun; Rwanda -Burundi).

When considering 25 candidate species (the 24 MOTUs identified in the ABGD analysis with the recursive partition and JC69 distance and *H. aeta* divided in two candidate species: Bioko - Mt Cameroun and Rwanda - Burundi), all analyses support the existence of 24 species (all specimens from *H. pamfi* are grouped into one candidate species) with high posterior probability, and the posterior probability of each delimited species is also high (PP > 0.95; Supplementary Data S8, Fig. 2).

3-5 Biogeographical analysis and divergence time analyses

Similar divergence dates are obtained for the three analyses (mito-nuclear supermatrix, nuclear and mitochondrial datasets), but confidence intervals are larger for the nuclear dataset (Supplementary Data S9). When we consider our *Hylomyscus* ingroup we identified 177 informative sites in the nuclear dataset, as compared to 3762 informative sites in the mitochondrial dataset. Because of the lack of nucleotide variations into the recent part of the tree in our nuclear dataset we think that the concatenation of both nuclear and mitochondrial genes is here necessary. Most of our signal is bear by the mitochondrial locus, which drives our molecular dating results but which will not be subject to lack of nucleotidic variations such as our nuclear dataset. Thus, in the rest of the manuscript we consider only the mito-nuclear results.

The results from the "DEC" and "DIVA-like" models were very similar and only differ in the ancestral areas and dispersal or vicariance events inferred for some nodes within the *parvus+anselli* group (Supplementary Data S10).

Both biogeographical analyses infer the continuous distribution of ancestral *Hylomyscus* in Late Miocene forests that covered eastern and central Africa, followed by a vicariance event (ca 5.7 Mya) that separated the east African *denniae* group from remaining *Hylomyscus* species (Fig. 3; Supplementary Data S11). The ancestor of the *baeri* and *aeta* groups probably lived in West and Central Africa, and a vicariance event separating the West African *H. baeri* from the Central African *aeta* group occurred at ca. 4.1 Mya. According to the DIVA-like model, at the same time period a dispersal event from Central to East Africa explains the widespread distribution of the ancestor of the *parvus* and *anselli* groups. It was followed by vicariance separating the Central African *parvus* group from the East African *anselli* group. In contrast, according to the DEC model (Supplementary Data S10) the ancestor of the *parvus* and *anselli* groups was restricted to Central Africa and a dispersal event from Central to East Africa occurred later (ca 3.6 Mya) and lead to the widespread distribution of the ancestor of the *anselli* group. According to both models the ancestor of the *anselli-kerbispeterhansi-sp7* clade was widely distributed in Central and East Africa, and a vicariance event restricting the two clades of *H. sp7* to Central Africa and dispersal of the *anselli-kerbispeterhansi* clade to East Africa occurred ca 1.3 Mya.

6- Discussion

6-1 Systematics, taxonomy and geographical distribution

In terms of gene sampling, geographical and taxonomic coverage, this study is the most complete phylogeny of the genus *Hylomyscus* to date. It is the first study to include representatives of all six species groups delimited by external and craniodental morphology. The five species groups *aeta*, *alleni*, *anselli*, *baeri* and *denniae* are monophyletic in both genetic datasets. The *parvus* group is recovered as monophyletic only in the mtDNA dataset, and with moderate support. The species groups *parvus* and *anselli* are sister taxa, and form a clade with *alleni*. The species groups *aeta* and *baeri* are sister clades, which is in agreement with morphological data showing that they both display a well-developed supraorbital shelf and a large and distinct cusp t9 on M^1 (Carleton et al., 2006). Finally, the *denniae* species group has a sister relationship to all other *Hylomyscus* species groups.

Based on the results of phylogenetic and species delimitation analyses we can review the systematics and taxonomy of each species group. In doing so, we consider species as separately evolving metapopulation lineages (de Queiroz, 2007). Because of incomplete lineage sorting, reciprocal monophyly is not necessarily a property of species, especially for recently diverged species (Knowles and Carstens, 2007). For diploid organisms, the effective population size of nuclear DNA is four times higher than that of the haploid and maternally inherited mitochondrial DNA. The faster coalescence of mitochondrial loci explains why a species can appear as monophyletic in the mitochondrial tree but not in the concatenated nuclear tree, in particular in the case of recent divergence (Leliaert et al., 2014). This is the case for the species *H. stella* and *H. walterverheyeni*, and *H. simus1* and *H. simus2* in our study. The multispecies coalescent method used in BPP allows testing of species boundaries in such shallow divergences (Yang and Rannala, 2010). In our review of each species group we consider as "species" those monophyletic groups in the mitochondrial DNA dataset that are also recovered in the BPP analysis based on nDNA (even if the monophyly is not supported in the nDNA tree).

alleni group (Table 1, Fig. 4A)

Within the *alleni* group, widely distributed in Western and Central Africa (North of the Congo River and along the Albertine Rift), our molecular data confirm the distinctiveness of the species *H. walterverheyeni* and *H. stella* which appear as sister species separated by the Oubangui-Congo river. *Hylomyscus stella kaimosae* is represented here by one virtual topotype from Kakamega Forest and we confirm that *kaimosae* is a synonym of *stella* (the specimens named *kaimosae* in the study of Nicolas et al. (2006; 2012), and shown to be genetically distinct from *H. stella*, were misidentified and corresponds to *H. arcimontensis*). No specimen from Bioko was included in the description of the species *H. walterverheyeni* (Nicolas et al., 2008b), but we found that several specimens from this island clustered genetically unambiguously with mainland specimens of *H. walterverheyeni*. We thus propose to extend its geographical distribution to Bioko Island.

mtDNA data and BPP analyses using nDNA genes support the recognition of two sister species in Cameroon and Bioko: *H.* cf. *alleni* and *H. sp2*. The species *H. alleni* has a complicated taxonomic history. It was described by Waterhouse (Waterhouse, 1837) from Bioko Island. Rosevear (1969) restricted *H. alleni* to Bioko, but subsequently several authors (e.g. Eisentraut, 1969; Musser and Carleton, 2005; Robbins et al., 1980) considered that it also occurs on the mainland of West-Central Africa. The fact that the holotype is a juvenile

specimen, with immature dentition, compromises the utility of the species description, and has lead several authors to consider this species as essentially unidentifiable (reviewed by Rosevear, 1969). An additional problem is that the exact type locality is unknown, even if, according to Rosevear (1969), it was probably "Clarence Bay" near Santa Isabel. Eisentraut (1966; 1969) reviewed the specimens from Bioko and Cameroon, gave detailed descriptions of their external and cranial morphology, and defined several topotypes. Pending molecular confirmation that the type specimen of *H. alleni* clusters with the specimens captured on the mainland, we propose that the specimens referred to *H.* cf *alleni* in our study correspond to true *H. alleni*. Should future analyses agree with this point of view, *H. alleni* is widely distributed in lowland forests from Bioko and Cameroon to central DRC (all records are from the right bank of the Congo River).

Two taxa are considered synonyms of *H. alleni* (Happold, 2013): *canus* described in 1940 from Eshobi (Western Cameroon, lowland forest) and *montis* described in 1969 from Pic van Santa Isabel (2000 m) in Bioko. The form *canus* was distinguished on the basis of its shorter, broader and stouter skull as compared to typical *alleni*. The form *montis* was described based on its longer ears, and shorter skull and rostrum compared to typical *alleni*. We sequenced the Cytochrome b gene of two specimens labeled *H. alleni montis* from the Koenig museum (ZFMK 69-661 and 66-733) and originating from the type locality of *montis*. They both clearly cluster with our *H. sp2*, suggesting that the name *montis* could be used for *H. sp2*. However, our specimens from Korup and Mt Kupé do not have long ears (mean = 14.4 mm, range = 13-16 mm). Comparison of the skull measurements of our specimens from Korup with the type specimens of *canus* and *montis* show that all these specimens are morphologically close to one another (Supplementary Data S13). Therefore, we cannot rule out the hypothesis that *montis* is a younger synonym of *canus*. Molecular analyses of the two type specimens are necessary for final determination. If future analyses confirm the synonymy of these two forms, then the older name (i.e. *canus*) should be used for *H. sp2*. Based on our results and those of Eisentraut (1969) *H. sp2* is present both in Bioko and Cameroon (Korup, Mt Kupe and Mt Oku), both in the lowlands and at high elevations.

Our genetic data confirm that all the specimens captured between the Volta and Niger Rivers belong to a distinct species: *H. pamfi* (Nicolas et al., 2010). Based on mtDNA data, *H. pamfi* is divided in two clades (specimens from Benin and from Nigeria, respectively), but this is not supported by nDNA data, and instead represents intraspecific phylogeographic structure.

H. simus was described as a subspecies of *H. alleni*, but was recently elevated to species rank based on mitochondrial DNA and morphometric analyses (Nicolas et al., 2010; Nicolas et al.,

2006). Our genetic data confirm its specific status, and suggest that it is composed of two separate gene pools, possibly representing two morphologically similar species with allopatric distributions: one (*H. simus1*) is distributed from W Guinea to W Ivory Coast, and the other one (*H. simus2*) is distributed from Ivory Coast to Ghana, up to the Volta River. *Hylomyscus simus* was described from Merikay (interior Liberia), within the distribution of *H. simus1*. If we accept the presence of two species, *H. simus2* would be an undescribed species, which needs formal description.

Based on mitochondrial genes and morphometric analyses Nicolas et al. (2012) and Kennis (2012) suggested that an additional species, genetically close to *H. walterverheyeni* and *H. stella*, occurs in Masako (DRC, right bank of the Congo River). This species was called "sp6" by Nicolas et al. (2012) and "taxon A" by Kennis (2012). These specimens were erroneously identified as *H. parvus* by Dudu et al. (1989), but they differ both genetically and morphologically from that species (Kennis, 2012). Unfortunately, this candidate species was not included in our study as complete mitochondrial DNA sequences and nDNA data could not be obtained from these specimens.

anselli group (Table 1, Fig. 4B)

Hylomyscus denniae anselli was described by Bishop (1979) based on its short incisive foramina and other characters. It was lumped within *H. denniae* by Musser and Carleton (2005). However based on their morphological revision of the *H. denniae* group, Carleton & Stanley (2005) elevated *H. anselli* to specific status and described a new species, *H arcimontensis*. According to these authors the former nominal *H. denniae* living in montane forests in Albertine Rift and eastern Africa was a composite of several species whose interrelationships and differentiation patterns indicate two species groups: the *denniae* group and the *anselli* group. Subsequent molecular work by Demos et al. (2014a; 2014b) confirmed these two species groups and defined a third species (*H. kerbispeterhansi* from montane Kenya) within the *anselli* group. Moreover, these two groups are not closely related. More recently, *H. heinrichorum* from the highlands of Angola was added to the *anselli* group (Carleton et al., 2015) resulting in four named species in this clade (*anselli, arcimontensis, kerbispeterhansi* and *heinrichorum*).

Our multilocus phylogenetic analysis infers two main genetic subclades within the *anselli* group: one corresponding to *H. heinrichorum* + *H. anselli* and the other corresponding to *H. kerbispeterhansi* + *H.* cf. *anselli* + *H. sp7* + *H. arcimontensis*. The geographic range of *H. anselli*, originally described from northwestern Zambia, was considered by Carleton and

Stanley (2005) to encompass Zambia and westernmost Tanzania. Our genetic data show that all specimens from western Tanzania mistakenly assigned to *anselli* by Carleton & Stanley (2005) and Carleton et al. (2015) belong to a new genetically very distinct undescribed species here labeled as *H.* cf. *anselli*. All but one species in the *anselli* group are restricted to the montane forests to the south and east of the Congo Basin. The only exception is *H. sp.7* found in the lowland Guineo-Congolese forest on the left bank of the Congo River. Substantial mitochondrial genetic variability is present in *H. sp.7* and two MOTUs are recognized in the BPP analyses based on nDNA data. These MOTUs exist in close proximity to each other and are separated by the Lomami River. According to Kennis (2012) they cannot be distinguished morphometrically (skull measurements) suggesting that these MOTUs could simply represent intra-specific phylogeographic structure. More detailed geographic sampling on the left bank of the Congo River is required to perform the genetic and morphometric analyses that are required to establish the taxonomic status of these two MOTUs.

parvus group (Table 1, Fig. 4C)

Hylomyscus parvus was described from Belinga primary lowland forest (Gabon) based on its small size, short round head and stout body, brown dorsal pelage, shorter maxillary toothrow and distinct dental morphology. This species occurs in West-Central Africa where it is known from just a few localities (Robbins et al., 1980).

Our mtDNA and nDNA data show high genetic variability within *H. parvus*, with three well differentiated groups, which likely represent three distinct species. *Hylomyscus parvus1* is present in Cameroon, CAR and northern Gabon (North of the Ogooue River); *H. parvus2* and *H. parvus3* co-occur in southern Gabon (South of the Ogooue River), and *H. parvus2* is also present in the north of the Congo Republic. Based on the geographic origin of the type specimen, the name *parvus* could belong to *H. parvus1* but this has to be confirmed by sequencing of the type specimen and/or morphometric comparison of the type specimen with sequenced specimens.

baeri group

The *baeri* group is monospecific and includes only *H. baeri*. This species was described half a century ago (Heim de Balsac and Aellen, 1965), but is still known by just a few specimens (Monadjem et al., 2015), and only specimens from Ziama forest (Guinea) were sequenced.

Considering its relatively wide distribution from Sierra Leone to the Volta River in Ghana and intensive collecting in this region in recent decades, it is likely to be a rare species.

aeta group (Table 1, Fig. 4D)

Our multilocus phylogenetic analysis revealed two main clades within the *aeta* group (the phylogenetic position of *H. grandis* from Mt. Oku in Cameroon remains unknown, because this species was not sampled). One corresponds to *H. aeta*, currently considered to be widely distributed on the right bank of the Congo River from Cameroon to Albertine Rift, and one corresponds to an undescribed species from Gabon called here *H. sp1* (sensu Nicolas et al., 2012). These two taxa also differ by skull morphology (Kennis, 2012) and number of nipples: *H. aeta* has 3 pairs of nipples (1 + 2 = 6) and *H. sp1* 4 (2 + 2 = 8). Given that these two taxa can be discriminated both on genetic and morphological ground we propose that *H. sp1* is a new species that will be described elsewhere.

H. aeta was described from Bitye (SE Cameroon) and is known to occur in Bioko and Central African forests, from Cameroon to Burundi. Based on our mtDNA analyses, high genetic variability is observed within *H. aeta* with three MOTUs: Bioko - Mt Cameroon, Burundi and Rwanda, the latter two being sister taxa. Two of these MOTUs are also recognized in the BPP analyses based on nDNA data: Bioko - Mt Cameroun, and Rwanda - Burundi. These results suggest that *H. aeta* could represent a complex of species, with one species in West Central Africa, and one species in East Africa. Comparison of our data with sequences from GenBank suggests that the latter taxon is also present in DRC, on the right bank of the Congo River (Bomane and Masako). Additional analyses are necessary to confirm whether these taxa represent distinct species or instead represent intraspecific phylogeographic structure, and to assess their taxonomy and geographical distribution. Three names are presently recognized as synonyms of *aeta*: *laticeps*, originally described based on a single specimen from high altitude forest of Mt Cameroon, *schoutedeni*, described based on two specimens from Mambaka (DRC), and *weileri*, described from Mt Mikeno (Rwanda, 2400 m).

According to Happold (2013) *H. aeta* also occurs on the left bank of the Congo River in DRC. However Kennis (2012) showed that DRC specimens from the left bank of the Congo River (Yoko and Djabir localities, named ‘taxon C” by Kennis) represent the sister clade of *H. aeta* (Cytochrome b gene sequencing). Moreover, specimens from the left bank of the Congo River also differ in their skull morphology from Cameroon and DRC right bank specimens. These results suggest the presence of a new undescribed species on the left bank of the Congo River, which is not surprising given that this river is known to be a geographic barrier for many

rodent species (Bryja et al., 2014b; Katuala et al., 2008; Kennis et al., 2011; Nicolas et al., 2008a; Olayemi et al., 2012).

denniae group (Table 1, Fig. 4E)

Hylomyscus denniae was described by Thomas (1906) from the Ruwenzori Mts. of western Uganda. Subsequently, *H. endorobae* (Heller, 1910) was described from the Kenyan Highlands and was promptly synonymized with *H. denniae* by Hollister (1919). *H. vulcanorum* was described as subspecies of *H. denniae* by Lönnberg & Gyldensolpe (1925), a designation accepted by Allen (1939). Musser and Carleton (2005) recognized *H. denniae* with *H. vulcanorum* and *H. endorobae* as synonyms. Based on genetic evidence, the monophyly and genetic differentiation of the three species in the *denniae* group (*denniae*, *endorobae*, *vulcanorum*) are now well established (Demos et al., 2014a; Demos et al., 2014b; Demos et al., 2015; Huhndorf et al., 2007), and confirmed by our study. Species in the *denniae* group are distributed in montane forests outside the Guineo-Congolese forest of the Congo Basin and are endemic to the Albertine Rift Mts and the Kenyan Highlands.
On the Mau Escarpment in Kenya, the *denniae* group overlaps with the *anselli* group, specifically *H. endorobae* and *H. kerbispeterhansi* were recorded in syntopy in montane forest (Demos et al., 2014a; Demos et al., 2014b). Even though Bishop (1979) mentioned several specimens of *Hylomyscus* from Ngorongoro and Tengeru in Tanzania, repeated recent small mammal surveys of the northern highlands of Tanzania (e.g. Mt Kilimanjaro, Mt Meru, Lake Manyara vicinity) and southern Kenya (Taita Hills) have failed to reveal the presence of *Hylomyscus* (our unpubl. data). Thus, a distributional gap in *Hylomyscus* likely occurs between the Eastern Arc Mts (*H. arcimontensis*) and the Kenyan Highlands (*H. endorobae*, *H. kerbispeterhansi*). Whether this gap is the result of regional extinction or a lack of dispersal between both areas is unknown.

6-2 Evolutionary history of the genus *Hylomyscus*

Our multilocus phylogeny supports the monophyly of the genus *Hylomyscus*, and a Central + East African origin of the genus. The first split within this genus occurred at the end of the Miocene and separated the *denniae* group (endemic to the Albertine Rift and Kenya Highlands montane forests of East Africa), from its sister lineage whose ancestor probably occurred in Central Africa. The formation of the Rift Valley combined with declining global temperatures during the Late Miocene caused greater precipitation seasonality, which resulted in the expansion of grasslands and forest fragmentation (Bobe, 2006). A similar split between

Central and East African species occurred at the end of the Miocene, as has been already observed in numerous other taxa, including rodents (Bryja et al., 2014a; Bryja et al., 2017), snakes (Portillo et al., 2018) and duikers (Johnston and Anthony, 2012). These findings support a scenario in which continuous African Miocene forest became divided into the current Central and West African Guinea-Congolese forest the East African forests (Plana, 2004).

The Early Pliocene (5-3.5 Mya) marked a return of a hotter and a more humid climate, representing the warmest period during the last 5 My. This has been termed the 'Golden Age', with tropical forests extending to 20°N (Morley, 2000; Plana, 2004). This expansion possibly reconnected West/Central with Eastern rain forest blocks. Indeed, fossils sites from East Africa document the presence of moist-adapted taxa and forest around 5-3 Mya (Jacobs, 2004; Linder, 2017; Morley, 2000; Pickford et al., 2004). Subsequent to the late Miocene and early Pliocene, the Congo Basin became humid again (Senut et al., 2009). The Turkana gap fossil site in southern Ethiopia dated to 3.4-3.3 Mya documents the presence of plant (*Antrocaryon,* Anacardiaceae) and animal (*Potadoma*, Pachychilidae) taxa known today only from Central African rain forests (Williamson, 1985). During this period, the resulting continuous forest cover facilitated two dispersal events within the genus *Hylomyscus*: one from Central to West Africa, and one from Central to East Africa. The early Pliocene was not only characterized by forest range extension, but also by two prolonged periods of high climate variability (Potts and Faith, 2015). These successive phases of increased wet-dry variability may have favored divergence events within *Hylomyscus*, as observed in other mammal taxa (Bryja et al., 2014b; Potts and Faith, 2015).

Intensive radiation within *Hylomyscus* is dated to the period between 3.5 and 1.4 Mya (end of the Pliocene, beginning of the Pleistocene), when most current species appeared. The onset of this period coincides with the start of a regional cooling and aridification trend (Plana, 2004). During this period of increased climate variation, open savannah-like habitats expanded significantly. According to Potts & Faith (2015) four prolonged periods of strong wet-dry variability occurred. This fluctuating climate probably led to environmental changes that increased the frequency and intensity of evolutionary processes like local adaptations, altered distribution ranges and ultimately facilitated speciation, as already documented in many other studies (Bryja et al., 2014a; Bryja et al., 2014b; Bryja et al., 2017; Demos et al., 2015; Huntley and Voelker, 2016; Potts and Faith, 2015; Voelker et al., 2010). Periods of increased climatic instability resulted in elevated diversification in both lowland forest taxa (see examples in the *parvus*, *alleni* and *aeta* groups) and montane forest species taxa (*denniae* and

anselli groups). The driest periods resulted in the fragmentation of lowland forests and the formation of lowland forest refugia, which favored allopatric speciation of lowland adapted taxa. At the same time, lower temperatures resulted in the expansion of montane and sub-montane vegetation to lower altitudes, resulting in the partial replacement of lowland rainforest, and facilitating the dispersal of montane floras and faunas between the previously isolated mountain ranges. With each intermittent period with a warmer and more humid climate, these montane forests and their faunas would retreat to higher altitudes, such that they would become effectively isolated, which would have promoted allopatric diversification, in contrast to widely distributed lowland forest taxa. In this context, it is remarkable that the closest relatives of the lowland taxon *H. sp7* are montane forest specialists. Other members of the *anselli* group are montane forest species, however, it should be acknowledged that additional sampling on the left bank on the Congo River may reveal new undescribed lowland species within this group (Kerbis Peterhans et al., in review). The historical biogeography of this species group was not fully resolved in our analyses (DIVA-like and DEC models gave different scenarios), and may be much more complex than presently understood, with multiple dispersals between montane to lowland and lowland to montane forests possible.

The last period of climatic instability is dated 1.1-0.9 Mya (Potts and Faith, 2015), and corresponds to the split between *H. anselli* (endemic to mid-altitude forests in Zambia) and *H. heinrichorum* (endemic to montane forest of Angola), and between the *H.* cf. *anselli* (montane forest of SW Tanzania) and *H. kerbispeterhansi* (montane forest of Kenya Highlands). This period also corresponds to the Early-Middle Pleistocene Transition (EMPT). The EMPT is characterized by the marked prolongation (glacial–interglacial cycles occur with a mean quasi periodicity of ~100 kyr after the EMPT) and intensification (dramatic increase of the contrast between warm and cold periods) of glacial-interglacial climate cycles (Maslin et al., 2014). After the EMPT these marked and prolonged glacial-interglacial cycles may have favored diversification within the taxa *H. sp7* (two clades separated by the Lomami River) and *H. aet*a (Cameroon to CAR; and D.R.C. to Rwanda-Burundi).

7- Conclusions

This study is the most complete phylogeny of the genus *Hylomyscus* to date. It supports the six species groups previously defined on morphological grounds as monophyletic, and reveals undescribed species diversity within this genus. As many as 10 taxa need descriptions or elevation from synonymy, pending review of type specimens. Integrative taxonomic studies

including type specimens are now required to confirm the names proposed here and to formally describe new species. Moreover, additional sampling on the left bank of the Congo River in Democratic Republic of Congo, as well as in Angola, may well lead to the discovery of more species. All divergence events within the genus occurred after the end of the Miocene and are tightly linked to the history of African forests. The formation of the Rift Valley combined with the declining global temperatures during the Late Miocene are associated with the fragmentation of pan-African rain forests and correlate with the split between the *denniae* group from sister clades. Subsequent periods of increased climatic instability during the Plio-Pleistocene resulted in elevated diversification in both lowland and montane *Hylomyscus* taxa, contributing to the high diversity present in this speciose rodent genus.

Acknowledgements

This work was supported by the 'Action Transversale du Muséum: Génomique et collections' (MNHN). All molecular lab work was performed at the 'Service de Systématique Moléculaire' of the MNHN (UMS 2700 2AD, Paris, France). Field campaigns were funded by EU-DGVIII BIOFAC and ECOFAC programs, WHO-Ebola Research Project in Côte d'Ivoire, WWF Gabon, PGRR-GFA Terra Systems Project in Guinea, INCO-DEV Threat Control in Guinea and Côte d'Ivoire, PAMF project in Benin, ANR Biodiversité IFORA ANR-06-BDIV-0014, the projects of the Czech Science Foundation nos. P506/10/0983 and 15-20229S, Grant Agency of the University of South Bohemia no. 018/2017/P, the Belgian Directorate for Development Cooperation (DGD), the Flemish Inter-University Council – University Development Cooperation (VLIR-UOS) We thank the Kenya Forest Service (KFS), Kenya Wildlife Service (KWS), and the National Museums of Kenya (NMK) for facilitating logistics that allowed small mammal surveys to be conducted in Kenya. We thank all African authorities for their assistance and permit authorizations. We are grateful to the curators of the following museums for allowing us to use their collections: FMNH, Chicago; CBGP, Montferrier-sur-Lez; University of Antwerp; IRSNB, Bruxelles; RMCA, Tervuren, ZFMK, Bonn. We are grateful to all collectors and particularly B. Agwanda, S. M. Goodman, G. Nikolaus, M. Solomon, M. H. Huhndorf, R. H. Pine, P. Kaleme, R. Šumbera, E. Lecompte, A. Lalis, A. Konečný, S. Šafarčíková, H. Konvičková, M. Lövy, J. Škliba, J. Krásová, O. Mikula, F. Vejmělka, V. Mazoch, G. Mbalitini, C. K. Kyamakya, G. G. Tungaluna, P. M. Kabeya, N. A. Draso, P. Musaba. We thank Agnès Dettai for her help in molecular lab work and preliminary data analyses.

Funding

Funding: This work was supported by the 'Action Transversale du Muséum: Génomique et collections' (MNHN), EU-DGVIII BIOFAC and ECOFAC programs, WHO-Ebola Research Project in Côte d'Ivoire, WWF Gabon, PGRR-GFA Terra Systems Project in Guinea, INCO-DEV Threat Control in Guinea and Côte d'Ivoire, PAMF project in Benin, ANR Biodiversité (grant number IFORA ANR-06-BDIV-0014), Czech Science Foundation (grant numbers P506/10/0983, 15-20229S), Grant Agency of the University of South Bohemia (grant number 018/2017/P), the Belgian Directorate for Development Cooperation (DGD), the Flemish Inter-University Council –University Development Cooperation (VLIR-UOS).

References

Aghova, 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). Mol Phylogenet Evol 128, 98-111.

Allen, G.M., 1939. A cheklist of African mammals. Bull Mus Comp Zool 83, 1-763.

Benton, M.J., Donoghue, P.C., 2007. Paleontological evidence to date the tree of life. Mol Biol Evol 24, 26-53.

Benton, M.J., Donoghue, P.C.J., Asher, R.J., 2009. Calibrating and constraining molecular clocks. In: Hedges, S.B., Kumar, S. (Eds.), The timetree of life. Oxford University Press, United Kingdom, pp. 35-86.

Bishop, I.R., 1979. Note on *Praomys* (*Hylomyscus*) in Eastern Africa. Mammalia 43, 519-530.

Bobe, R., 2006. The evolution of arid ecosystems in eastern Africa. J Arid Environ 66, 564-584.

Bryja, J., Mazoch, V., Patzenhauerova, H., Mateke, C., Zima, J.J., Skliba, J., Sumbera, 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., Patzenhauerov, H., Oguge, N.O., Sumbera, R., Verheyen, E., 2014a. The role of dispersal and vicariance in the Pleistocene history of an East African mountain rodent, *Praomys delectorum*. J Biogeogr 41, 196-208.

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

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

Carleton, M.D., Banasiak, R.A., Stanley, W.T., 2015. A new species of the rodent genus *Hylomyscus* from Angola, with a distributional summary of the *H. anselli* species group (Muridae: Murinae: Praomyini). Zootaxa 4040, 101-128.

Carleton, M.D., Kerbis Peterhans, J.C., Stanley, W.T., 2006. Review of the *Hylomyscus denniae* group (Rodentia: Muridae) in Eastern Africa, with comments on the generic allocation of *Epimys endorobae* Heller. Proc Bio Soc Wash 119, 293.

Carleton, M.D., Stanley, W.T., 2005. Review of the *Hylomyscus denniae* complex (Rodentia: Muridae) in Tanzania, with a description of a new species. Proc Bio Soc Wash 118, 619-646.

Coulibaly-N'Golo, D., Allali, B., Kouassi, S.K., Fichet-Calvet, E., Becker-Ziaja, B., Rieger, T., Olschlager, S., Dosso, H., Denys, C., Ter Meulen, J., Akoua-Koffi, C., Gunther, S., 2011. Novel arenavirus sequences in *Hylomyscus* sp. and *Mus* (*Nannomys*) *setulosu*s from Cote d'Ivoire: implications for evolution of arenaviruses in Africa. PLoS ONE 6, e20893.

de Queiroz, K., 2007. Species concepts and species delimitation. Systematic Biol 56, 879-886.

Demos, T.C., Agwanda, B., Hickerson, M., 2014a. Integrative taxonomy within the *Hylomyscus denniae* complex (Rodentia: Muridae) and a new species from Kenya. J Mammal 95, E1-E15.

Demos, T.C., Kerbis Peterhans, J.C., Agwanda, B., Hickerson, M.J., 2014b. Uncovering cryptic diversity and refugial persistence among small mammal lineages across the Eastern Afromontane biodiversity hotspot. Mol Phylogenet Evol 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(9), e0131800.

Denys, C., Missoup, A.D., Tchiengue, B., Achoundong, G., Ekobo, A., Bilong Bilong, C.F., Lembe, D.M., Nicolas, V., 2009. Altitudinal distribution of and anthropogenic influence on small mammal assemblages on Mount Kupe, SW Cameroon. Bonn Zool Beitr 56, 159-173.

Dudu, A.M., Van der straeten, E., Verheyen, W.N., 1989. Première capture de *Hylomyscus parvus* Brosset, Dubost et Heim de Balzac, 1965 au Zaïre avec quelques données biométriques (Rodentia, Muridae). Rev Zool Afr 103, 179-182.

Duplantier, J.M., 1989. Les rongeurs myomorphes forestiers du nord-est du Gabon: structure du peuplement, démographie, domaines vitaux. Rev Ecol Terre Vie 44, 329-346.

Eisentraut, M., 1966. Die *Hylomyscus*-Formen von Fernando Poo. Z Säugetierkunde 31, 213-219.

Eisentraut, M., 1969. Die verbreitung der Muriden-Gattung *Hylomyscus* auf Fernando Poo und in Westkamerun. Z Säugetierkunde 31, 296-307.

Fayolle, A., Swaine, M.D., Bastin, J.-F., Bourland, N., Comiskey, J.A., Dauby, G., Doucet, J.-L., Gillet, F.-F., Gourlet-Fleury, S., Hardy, O., Kirunda, B., Kouame, F.N., Plumptre, A.J., 2014. Patterns of tree species composition across tropical African forests. J Biogeog 41, 2320-2331.

Flot, J.F., 2010. Seqphase: a web tool for interconverting phase input/output files and fasta sequence alignments. Mol Ecol Res 10, 162-166.

Happold, D., 2013. Mammals of Africa, Volume III: Rodents, Hares and Rabbits. Bloomsbury Publishing, London.

Heim de Balsac, H., Aellen, V., 1965. Un *Hylomyscus* nouveau de Côte d'Ivoire. Biol Gabonica 1, 175-178.

Heller, E., 1910. Descriptions of seven new species of east African mammals. Smithson Miscel Coll 56, 1-5.

Hinsinger, D.D., Debruyne, R., Thomas, M., Denys, G.P.J., Mennesson, M., Utge, J., Dettaï, A., 2016. Fishing for barcodes in the Torrent: from COI to complete mitogenomes on NGS platforms. DNA Barcodes 3, 170-186.

Hollister, N., 1919. East African mammals in the United States National Museum. Part II. Rodentia, Lagomorpha, and Tubulidentata. US Nat Mus Bull 99, 1-184.

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.

Huntley, J.W., Voelker, G., 2016. Cryptic diversity in Afro-tropical lowland forests: The systematics and biogeography of the avian genus *Bleda*. Mol Phylogenet Evol 99, 297-308.

Igea, J., Juste, J., Castresana, J., 2010. Novel intron markers to study the phylogeny of closely related mammalian species. BMC Evol Biol 10, 369.

Jacobs, B.F., 2004. Paleobotanical studies from tropical Africa: relevance to the evolution of forest, woodland and savannah biomes. Philos Trans R Soc B 359, 1573-1583.

Jacobs, L.L., Flynn, L.J., 2005. Of Mice…again: the Siwalik rodent record, murine distribution, and molecular clocks. In: Lieberman, D.E., Smith, R.J., Keller, J. (Eds.), Interpreting the Past Essays on Human, Primate, and Mammal Evolution in Honor of David Pilbeam. Brill Academic Publishers, pp. 63–80.

Johnston, A.R., Anthony, N.M., 2012. A multilocus species phylogeny of African forest duikers in the subfamily Cephalophinae: evidence for a recent radiation in the Pleistocene. BMC Evol Biol 12, 120.

Katuala, P.G.-B., Kennis, J., Nicolas, V., Wendelen, W., Huselmans, J., Verheyen, E., Van Houtte, N., Dierckx, T., Dudu, A.M., 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.

Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., Buxton, S., Cooper, A., Markowitz, S., Duran, C., Thierer, T., Ashton, B., Meintjes, P., Drummond, A., 2012. Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics 28, 1647-1649.

Kennis, J., 2012. Study of tropical rainforest rodents: biodiversity and adaptation ability to changing habitats around Kisangani, D.R.Congo. Departement of Biology. University of Antwerp, Antwerpen, p. 179.

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., Hutterer, R., Doty, J.B., Malekane, J.M., Moyer, D.C., Demos, T.C., in review. Four new species of the *Hylomyscus anselli* group (Mammalia, Rodentia) from the Democratic Republic of Congo and Tanzania. Bonn Zool Bull.

Kimura, Y., Hawkins, M.T., McDonough, M.M., Jacobs, L.L., Flynn, L.J., 2015. Corrected placement of *Mus-Rattus* fossil calibration forces precision in the molecular tree of rodents. Sci Rep 5, 14444.

Kimura, Y., Jacobs, L.L., Cerling, T.E., Uno, K.T., Ferguson, K.M., Flynn, L.J., Patnaik, R., 2013. Fossil mice and rats show isotopic evidence of niche partitioning and change in dental ecomorphology related to dietary shift in Late Miocene of Pakistan. PLoS One 8, e69308.

Klempa, B., Fichet-Calvet, E., Lecompte, E., Auste, B., Aniskin, V., Meisel, H., Denys, C., Koivogui, L., ter Meulen, J., Kruger, D.H., 2006. Hantavirus in African wood mouse, Guinea. Emerg Infect Dis 12, 838-840.

Knowles, L.L., Carstens, B.C., 2007. Delimiting species without monophyletic gene trees. Syst Biol 56, 887-895.

Lanfear, R., Frandsen, P.B., Wright, A.M., Senfeld, T., Calcott, B., 2017. PartitionFinder 2: New Methods for Selecting Partitioned Models of Evolution for Molecular and Morphological Phylogenetic Analyses. Mol Biol Evol 34, 772-773.

Lartillot, N., Lepage, T., Blanquart, S. 2009. PhyloBayes 3: a Bayesian software package for phylogenetic reconstruction and molecular dating. Bioinformatics 25:2286–2288.

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., Granjon, L., Denys, C., 2002a. The phylogeny of the *Praomys* complex (Rodentia : Muridae) and its phylogeographic implications. J Zool Syst Evol Res 40, 8-25.

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

Leliaert, F., Verbruggen, H., Vanormelingen, P., Steen, F., López-Bautista, J.M., Zuccarello, G.C., De Clerc, O., 2014. DNA-based species delimitation in algae. European Journal of Phycology 49, 179-196.

Lemmon, A.R., Brown, J.M., Stanger-Hall, K., Lemmon, E.M., 2009. The effect of ambiguous data on phylogenetic estimates obtained by maximum likelihood and Bayesian inference. Syst Biol 58, 130-145.

Lepage, T., Bryant, D., Philippe, H., Lartillot N., 2007. A general comparison of relaxed molecular clock models. Mol Biol Evol 24, 2669-2680.

Linder, H.P., 2017. East African Cenozoic vegetation history. Evol Anthropol 26, 300-312.

Lönnberg, E., Gyldenstolpe, N., 1925. Zoological results of the swedish expedition to central Africa 1921 - Vertebrata - 2 - Preliminary diagnoses of seven new mammals. Arkiv för Zoologi 17B5, 1-6.

Maslin, M.A., Brierleya, C.M., Milner, A.M., Shultz, S., Trauth, M.H., Wilson, K.E., 2014. East African climate pulses and early human evolution. Quat Sc Rev 101, 1-17.

Matzke, N., 2013. BioGeoBEARS: BioGeography with Bayesian (and Likelihood) Evolutionary Analysis in R Scripts. University of California, Berkeley.

Missoup, A.D., Nicolas, V., Wendelen, W., Bilong Bilong, C.F., Cruaud, C., Denys, C., 2009. First record of *Hylomyscus walterverheyeni* on the north-western side of the Sanaga River (western Cameroon). Zootaxa 2044, 46-60.

Monadjem, A., Taylor, P., Denys, C., Cotterill, F.D.P., 2015. Rodents of Sub-Saharan Africa. A biogeographic and taxonomic synthesis. De Gruyter, Berlin, Boston.

Morley, R.J., 2000. Origin and evolution of tropical rain forests. Wiley J. and Sons Ltd, Chicheste.

Musser, G.G., Carleton, M.D., 2005. Superfamily Muroidea. In: Wilson, D.E., Reeder, D.M. (Eds.), Mammal species of the world: a taxonomic and geographic reference. The Johns Hopkins University Press, Baltimore, pp. 894-1531.

Nicolas, V., Colyn, M., 2003. Seasonal variations in population and community structure of small rodents in a tropical forest of Gabon. Can J Zool 81, 1034-1046.

Nicolas, V., Mboumba, J.F., Verheyen, E., Denys, C., Lecompte, E., Olayemi, A., Colyn, M., 2008a. Phylogeographical structure and regional history of *Lemniscomys striatus* (Rodentia: Muridae) in tropical Africa. J Biogeogr 35, 2072-2089.

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.F., 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., Wendelen, W., Barrière, P., Dudu, A., Colyn, M., 2008b. Morphometrical variation in *Hylomyscus alleni* and *Hylomyscus stella* (Rodentia, Muridae), and description of a new species. J Mammal 89, 222-231.

Olayemi, A., Akinpelu, A., 2008. Diversity and distribution of murid rodent populations between forest and derived savanna sites within south western Nigeria. Biodivers Conserv DOI 10.1007/s10531-008-9389-1.

Olayemi, A., Cadar, D., Magassouba, N., Obadare, A., Kourouma, F., Oyeyiola, A., Fasogbon, S., Igbokwe, J., Rieger, T., Bockholt, S., Jerome, H., Schmidt-Chanasit, J., Garigliany, M., Lorenzen, S., Igbahenah, F., Fichet, J.N., Ortsega, D., Omilabu, S., Gunther, S., Fichet-Calvet, E., 2016. New Hosts of The Lassa Virus. Sci Rep 6, 25280.

Olayemi, A., Nicolas, V., Hulselmans, J., Missoup, A.D., Fichet-Calvet, E., Amundala, D., Dudu, A., Dierckx, T., Wendelen, W., Leirs, H., Verheyen, E., 2012. Taxonomy of the African Giant Pouched Rats (Nesomyidae: *Cricetomys*): molecular and craniometric evidence support an unexpected high species diversity. Zool J Linn Society 165, 700–719.

Olteanu, M., Nicolas, V., Schaeffer, B., Denys, C., Missoup, A.D., Kennis, J., Laredo, C., 2013. Nonlinear projection methods for visualizing Barcode data and application on two data sets. Mol Ecol Resour 13, 976-990.

Padial, J.M., De la Riva, I., 2010. A response to recent proposals for integrative taxonomy. Biol J Linn Soc 101, 747-756.

Parham, J.F., Donoghue, P.C., Bell, C.J., Calway, T.D., Head, J.J., Holroyd, P.A., Inoue, J.G., Irmis, R.B., Joyce, W.G., Ksepka, D.T., Patane, J.S., Smith, N.D., Tarver, J.E., van Tuinen, M., Yang, Z., Angielczyk, K.D., Greenwood, J.M., Hipsley, C.A., Jacobs, L., Makovicky, P.J., Muller, J., Smith, K.T., Theodor, J.M., Warnock, R.C., Benton, M.J., 2012. Best practices for justifying fossil calibrations. Syst Biol 61, 346-359.

Philippe, H., Zhou, Y., Brinkmann, H., Rodrigue, N., Delsuc, F. (2005a). Heterotachy and long-branch attraction in phylogenetics. BMC Evol Biol 5, 50.

Philippe, H., Lartillot, N., Brinkmann, H. (2005b). Multigene analyses of bilaterian animals corroborate the monophyly of Ecdysozoa, Lophotrochozoa, and Protostomia. Mol Biol Evol 22, 1246-1253.

Pickford, M., Senut, B., Mourer Chauvire, U., 2004. Early Pliocene tragulidae and peafowls in the rift valley, Kenya: Evidence for rainforest in east Africa. Comptes Rendus Palevol 3, 179-189.

Plana, V., 2004. Mechanisms and tempo of evolution in the African Guineo–Congolian rainforest. Phil Trans R Soc Lond B 359, 1585-1594.

Portillo, F., Branch, W.R., Conradie, W., Rodel, M.O., Penner, J., Barej, M.F., Kusamba, C., Muninga, W.M., Aristote, M.M., Bauer, A.M., Trape, J.F., Nagy, Z.T., Carlino, P., Pauwels, O.S.G., Menegon, M., Burger, M., Mazuch, T., Jackson, K., Hughes, D.F., Behangana, M., Zassi-Boulou, A.G., Greenbaum, E., 2018. Phylogeny and biogeography of the African burrowing snake subfamily Aparallactinae (Squamata: Lamprophiidae). Mol Phylogenet Evol 127, 288-303.

Potts, R., Faith, J.T., 2015. Alternating high and low climate variability: The context of natural selection and speciation in Plio-Pleistocene hominin evolution. J Hum Evol 87, 5-20.

Puillandre, N., Lambert, A., Brouillet, S., Achaz, G., 2011. ABGD, Automatic Barcode Gap Discovery for primary species delimitation. Mol Ecol 21, 1864-1877.

Kumar, S., Stecher, G., Li, M., Knyaz, C., Tamura K., 2018. MEGA X: Molecular Evolutionary Genetics Analysis across Computing Platforms. Mol Biol Evol 35, 1547-1549.

Rambaut, A., Suchard, M.A., Xie, D., Drummond, A.J., 2013. Tracer v1.5. Available from http://beast.bio.ed.ac.uk/Tracer

Ree, R.H., Sanmartin, I., 2018. Conceptual and statistical problems with the DEC+J model of founder-event speciation and its comparison with DEC via model selection. J Biogeog 45, 741-749.

Ritchie, A.M., Lo, N., Ho, S.Y.W., 2017. The Impact of the Tree Prior on Molecular Dating of Data Sets Containing a Mixture of Inter- and Intraspecies Sampling. Syst Biol 66, 413-425.

Robbins, L.W., Choate, J.R., Robbins, R.L., 1980. Nongeographic and interspecific variation in four species of *Hylomyscus* (Rodentia: Muridae) in Southern Cameroon. Ann Carnegie Mus 49, 31-48.

Ronquist, F., Teslenko, M., van der Mark, P., Ayres, D.L., Darling, A., Hohna, S., Larget, B., Liu, L., Suchard, M.A., Huelsenbeck, J.P., 2012. MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. Syst Biol 61, 539-542.

Rosevear, D., 1969. The Rodents of West Africa. Trustees of the British Museum (Natural History), London.

Rosevear, D.R., 1966. The subgenus *Hylomyscus* Thomas in West Africa. Mammalia 30, 13-24.

Salicini, I., Ibanez, C., Juste, J., 2011. Multilocus phylogeny and species delimitation within the Natterer's bat species complex in the Western Palearctic. Mol Phylogenet Evol 61, 888-898.

Schenk, J.J., Rowe, K.C., Steppan, S.J., 2013. Ecological Opportunity and Incumbency in the Diversification of Repeated Continental Colonizations by Muroid Rodents. Syst biol 62, 837-864.

Srivathsan, A., Meier, R., 2012. On the inappropriate use of Kimura-2-parameter (K2P) divergences in the DNA-barcoding literature. Cladistics 28, 190-194.

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

Stephens, M., Smith, N.J., Donnelly, P., 2001. A new statistical method for haplotype reconstruction from population data. Am J Hum Genet 68, 978-989.

Steppan, S.J., Schenk, J.J., 2017. Muroid rodent phylogenetics: 900-species tree reveals increasing diversification rates. PLoS One 12, e0183070.

Sudi, L., Mangu, C., Tarimo, T., Ntinginya, N., Shirima, G., 2018. Hantaviruses in East and Central Africa. Am J Res Comm 6, 1-14.

Tesikova, J., Bryjova, A., Bryja, J., Lavrenchenko, L.A., Gouy de Bellocq, J., 2017. Hantavirus Strains in East Africa Related to Western African Hantaviruses. Vector Borne Zoonotic Dis 17, 278-280.

Thomas, O., 1906. Descriptions of new mammals from Mount Ruwenzori. Ann Mag Nat Hist, ser 7 18, 144-145.

Thorne, J.L., Kishino, H., Painter I.S., 1998. Estimating the rate of evolution of the rate of molecular evolution. Mol Biol Evol 15, 1647-1657.

Thorne, J.L., Kishino, H., 2002. Divergence time and evolutionary rate estimation with multilocus data. Syst. Biol. 51, 689-702.

Tilak, M.K., Justy, F., Debiais-Thibaud, M., Botero-Castro, F., Delsuc, F., Douzery, E.J.P., 2015. A cost-effective straightforward protocol for shotgun Illumina libraries designed to assemble complete mitogenomes from non-model species. Conservation Genet Resour 7, 37-40.

Voelker, G., Outlaw, R.K., Bowie, R.C.K., 2010. Pliocene forest dynamics as a primary driver of African bird speciation. Global Ecol Biogeogr 19, 111-121.

Waterhouse, G.R., 1837. Several small quadrupeds. Characters of some new species of the genera *Mus* and *Phascogale*. Proc. Zool. Soc. Lond, 75-78.

Williamson, P.G., 1985. Evidence for an early Plio-Pleistocene rainforest expansion in East Africa. Nature 315, 487-489.

Wilson, D.E., Lacher, T.E., Mittermeier, R.A., 2017. Handbook of the Mammals of the World. Vol 7 Rodents II. Lynx Editions, Barcelona.

Winnepenninckx, B., Backeljau, T., De Wachter, R., 1993. Extraction of high molecular weight DNA from molluscs. Trends Genet 9, 407.

Yang, Z., 2015. The BPP program for species tree estimation and species delimitation Current Zoology 61, 854-865.

Yang, Z., Rannala, B., 2010. Bayesian species delimitation using multilocus sequence data. Proc Natl Acad Sci U S A 107, 9264-9269.

Yu, Y., Harris, A.J., Blair, C., He, X., 2015. RASP (Reconstruct Ancestral State in Phylogenies): a tool for historical biogeography. Mol Phylogenet Evol 87, 46-49.

Figure legends:

Fig. 1: Phylogenetic tree estimated from Maximum Likelihood analysis of complete mitochondrial sequences. Circles at nodes represent bootstrap support (Maximum Likelihood analysis) and posterior probabilities (Bayesian analysis). The names of the species groups are in capital letters. The bars on the right side indicate the MOTUs found in ABGD analyses. For specimen numbers labels see Supplementary Data S4.

Fig. 2: Phylogenetic tree estimated from Maximum Likelihood analysis of nuclear DNA sequences. Circles at nodes represent bootstrap support (Maximum Likelihood analysis) and posterior probabilities (Bayesian analysis). The names of the species groups are in capital letters. The bars on the right side indicate the species supported by the BPP analyses (PP > 0.95) for 24, 40 and 25 candidate species, respectively. For specimen numbers labels see Supplementary Data S6.

Fig. 3: Chronogram of the genus *Hylomyscus* based on the mito-nuclear supermatrix with ancestral areas estimations from the "DIVA-like" biogeography analysis. To improve clarity, outgroup taxa are not represented (but see Supplementary Data S10 for the tree including outgroup taxa and 95% highest posterior density age ranges). Terminal taxa are color coded according to the biogeographical areas in which they occur. Colors at nodes indicate most likely ancestral areas recovered. Dispersion and vicariance events are indicated at nodes by the letters D and V, respectively. The tree displayed here is a chronogram based on PhyloBayes analysis of the combined dataset (mitochondrial DNA + 5 nuclear genes).

Fig. 4: Species distribution per species group: 4A = *alleni* group, 4B = *anselli* group, 4C = *parvus* group, 4D = *aeta* group and 4E = *denniae* group. For the localization of maps 4A to 4E on the African continent map see Supplementary Data S12. The species group *baeri* is not represented because it includes only one species from West Africa and specimens from only one locality (Ziama forest in Guinea) were sequenced. These maps are based on specimens for which the species identification was confirmed by genetic (GenBank data and our own unpublished data) or morphometric data (based on the studies of Bryja et al. (2012), Carleton et al. (2006, 2015), Demos et al. (2014a, 2014b), Denys et al. (2009), Happold & Happold (2013), Huhndorf et al. (2007), Kennis (2012), Missoup et al. (2009), Nicolas et al. (2008, 2013), Olteanu et al. (2013)). When several species co-occur in a given locality this is indicated by a pie chart.

Tables

Table 1:

Summary of the taxonomy of the genus *Hylomyscus* with geographical distribution and habitat requirement of all candidate species. Data based on Bryja et al. (2012), Carleton et al. (2015; 2006), Denys et al. (2009), Demos et al. (2014a; 2014b), Happold & Happold (2013), Huhndorf et al. (2007), Kennis (2012), Missoup et al. (2009), Monadjem et al. (2015), Musser & Carleton (2005), Nicolas et al. (2010; 2006; 2012; 2008b), Olteanu et al. (2013), Wilson et al. (2017) and our own unpublished data. DRC = Democratic Republic of Congo, CAR = Central African Republic. The * symbol in the column "proposed name" indicates that additional work is needed for confirmation. The species highlighted in grey could not be included in the present study.

Appendices

Appendix A. Supplementary material

Supplementary Data S1: List of specimens included in this study along with pertinent information (species name, locality of collect, GenBank numbers).

Supplementary Data S2: List of primers used to amplify the mitogenome by long PCRs.

Supplementary Data S3: Partition schemes inferred by PartitionFinder for the mtDNA dataset and the nDNA dataset.

Supplementary Data S4: Phylogenetic tree estimated from Maximum Likelihood analysis of complete mitochondrial sequences. The names of the species groups are indicated by a 2-letters code: AE = *aeta*, AL = *alleni*, AN = *anselli*, BA = *baeri*, DE = *denniae*, PA = *parvus*. FigS5A: Zoom on the *alleni* group; Fig. S5B: Zoom on the *anselli* and *parvus* groups; Fig. S5C: Zoom on the *aeta* and *baeri* groups; Fig. S5D: Zoom on the *denniae* group.

Supplementary Data S5: mean number of base substitutions per site (p-distance) per species.

Supplementary Data S6: Phylogenetic tree estimated from Maximum Likelihood analysis of nuclear DBA sequences. The names of the species groups are indicated by a 2-letters code: AE = *aeta*, AL = *alleni*, AN = *anselli*, BA = *baeri*, DE = *denniae*, PA = *parvus*. FigS7A: Zoom on the *alleni* group; Fig. S7B: Zoom on the *anselli* and *parvus* groups; Fig. S7C: Zoom on the *aeta* and *baeri* groups; Fig. S7D: Zoom on the *denniae* group.

Supplementary Data S7: ABGD results, with the number of MOTUs obtained for each prior intraspecific divergence for both the initial and recursive partitions.

Supplementary Data S8: Results of BPP species delimitation (A11 model) for algorithms 0 and 1. Analyses were performed on either 24, 25 or 40 candidate species on the nuclear dataset.

Supplementary Data S9: Molecular dating results obtained with PhyloBayes from the concatenated mitochondrial-nuclear dataset (SPM), and the nuclear or mitochondrial only datasets.

Supplementary Data S10: Chronogram derived from an analysis in PhyloBayes on the combined dataset (mitochondrial DNA + 5 nuclear genes). Node bars indicate 95% highest posterior density (HPD) age ranges. The three calibration points are highlighted with stars.

Supplementary Data S11: Fig. 3: Chronogram of the genus *Hylomyscus* based on the mito-nuclear supermatrix with ancestral areas estimations from the "DIVA-like" (left) and "DEC" (right) biogeography analysis. Terminal taxa are color coded according to the biogeographical areas in which they occur. Colors at nodes indicate most likely ancestral areas recovered. Dispersion and vicariance events are indicated at nodes by the letters D and V, respectively. The tree displayed here is the chronogram based on PhyloBayes analysis on the combined dataset (mitochondrial DNA + 5 nuclear genes).

Supplementary Data S12: map of the African continent showing the localization of maps 4A-4E of Fig. 4.

Supplementary Data S13: Morphometric comparison of *H. sp2* and the type specimens of *montis* and *canus*. HB: head and body length, TL: tail length, HF: hind-foot length, EL: ear length; GRLS: greatest length of skull, PRCP: condylobasal length, HEBA: henselion–basion length, HEPA: henselion–palation length, PAFL: length ofpalatal foramen, DIA1: length of diastema, DIA2: distance between M1 alveolus and cutting edge of upper incisor, INTE: smallest interorbital breadth, ZYGO: zygomatic breadth on the zygomatic process of the squamosal, PALA: smallest palatal breadth between 1st upper molars, UPTE: length of maxillary cheek toothrow, UPDA: breadth of upper dental arch (UPDA), M1BR: greatest breadth of M1, ZYPL: smallest breadth of zygomatic plate, BNAS: greatest breadth of nasals, LNAS: greatest length of nasals, LOTE: length of mandibular toothrow, BULL: length of auditory bulla, BRCA: greatest breadth of braincase, DINC: depth of I1, ROHE: mediosagittal projection of rostrum height at anterior border of 1st upper molars, ROBR: greatest rostrum breadth.

Highlights

- Mitochondrial genomes and nuclear genes are used to infer the evolutionary history of the genus *Hylomyscus*
- The six species groups defined on morphological grounds are monophyletic
- There are a high number of cryptic species
- Diversification after the end of the Miocene tightly linked to the history of the African forest

Author contribution statement

Nicolas Violaine: Conceptualization, Formal analysis, Investigation, Resources, Writing – original draft, Funding acquisition. **Fabre Pierre-Henri**: Formal analysis, Resources; Writing – original draft. **Bryja Josef**: Resources; Writing – review & editing. **Denys Christiane**: Resources. **Verheyen Erik**: Resources; Writing – review & editing. **Missoup Alain-Didier**: Resources. **Olayemi Ayodeji**: Resources. **Katuala Pionus**: Resources. **Dudu Akaibe**: Resources. **Colyn Marc**: Resources. **Kerbis Peterhans Julian**: Resources; Writing – review & editing. **Demos Terrence: Resources**; Writing – original draft.