Molecular Phylogenetics and Evolution 189 (2023) 107913

Contents lists available at ScienceDirect

Molecular Phylogenetics and Evolution

journal homepage: www.elsevier.com/locate/ympev

Unraveling phylogenetic relationships and species boundaries in the arid adapted *Gerbillus* rodents (Muridae: Gerbillinae) by RAD-seq data

Marcin Piwczyński [a,*], Laurent Granjon [b], Paulina Trzeciak [a], José Carlos Brito [c,d,e], Madalina Oana Popa [a,f], Mergi Daba Dinka [a], Nikolas P. Johnston [g,h], Zbyszek Boratyński [c,d]

[a] *Department of Ecology and Biogeography, Nicolaus Copernicus University in Toruń, Lwowska 1, PL-87-100 Toruń, Poland*
[b] *CBGP, IRD, CIRAD, INRAE, Institut Agro, Université de Montpellier, Montpellier, France*
[c] *CIBIO-InBio, Research Center in Biodiversity and Genetic Resources, University of Porto, Campus de Vairão, Rua Padre Armando Quintas 7, 4485-661 Vairão, Portugal*
[d] *BIOPOLIS Program in Genomics, Biodiversity and Land Planning, CIBIO, Campus de Vairão, Vairão, Portugal*
[e] *Departamento de Biologia, Faculdade de Ciências, Universidade do Porto, Porto, Portugal*
[f] *"Stejarul" Research Centre for Biological Sciences, National Institute of Research and Development for Biological Sciences, Alexandru cel Bun 6, RO-610004, Piatra Neamț, Romania*
[g] *School of Life Sciences, University of Technology Sydney, 15 Broadway, Ultimo, NSW 2007, Australia*
[h] *Centre for Sustainable Ecosystem Solutions, School of Earth, Atmospheric and Life Sciences, University of Wollongong, Northfields Ave, Wollongong, NSW 2500, Australia*

ARTICLE INFO

Keywords:
Desert
Molecular dating
Rodentia
Sahara-Sahel region
Species delimitation

ABSTRACT

Gerbillus is one of the most speciose genera among rodents, with ca. 51 recognized species. Previous attempts to reconstruct the evolutionary history of *Gerbillus* mainly relied on the mitochondrial cyt-b marker as a source of phylogenetic information. In this study, we utilize RAD-seq genomic data from 37 specimens representing 11 species to reconstruct the phylogenetic tree for *Gerbillus*, applying concatenation and coalescence methods. We identified four highly supported clades corresponding to the traditionally recognized subgenera: *Dipodillus*, *Gerbillus*, *Hendecapleura* and *Monodia*. Only two uncertain branches were detected in the resulting trees, with one leading to diversification of the main lineages in the genus, recognized by quartet sampling analysis as uncertain due to possible introgression. We also examined species boundaries for four pairs of sister taxa, including potentially new species from Morocco, using SNAPP. The results strongly supported a speciation model in which all taxa are treated as separate species. The dating analyses confirmed the Plio-Pleistocene diversification of the genus, with the uncertain branch coinciding with the beginning of aridification of the Sahara at the the Plio-Pleistocene boundary. This study aligns well with the earlier analyses based on the cyt-b marker, reaffirming its suitability as an adequate marker for estimating genetic diversity in *Gerbillus*.

1. Introduction

Deserts and arid regions (aridity index < 0.20; Ward, 2009) cover nearly 20% of the Earth's non-sea surface (Zomer et al., 2022). These areas host approximately one-fourth of the world's land vertebrate species, and are characterized by a high level of endemism of organisms adapted to extreme climatic conditions (Brito et al., 2016). Nonetheless, deserts and semi-deserts remain among the least scientifically investigated regions in the world (Durant et al., 2014).

The Sahara-Sahel ecoregions of Africa host a disproportionally high number of endemic species, resulting from radiation events in response to the onset and development of extreme and fluctuating arid conditions (Brito et al., 2014; Pausata et al., 2020). However, when it comes to understanding the evolutionary and ecological processes that have facilitated the emergence of these endemic species, the Sahara-Sahel ecoregions of Africa often present knowledge gaps. Such information would be essential for developing, among other topics, conservation strategies adapted to the potentially negative effects of climatic changes (Brito et al., 2021, 2014).

Here, we study *Gerbillus* rodents, which have been hypothesized to evolve and diversify in response to changing aridity conditions in North Africa, and subsequently colonize deserts and semi-deserts of Asia (Abiadh et al., 2010; Ndiaye et al., 2016a, 2016b). *Gerbillus* is one of the most speciose genera among rodents, with ca. 51 currently recognized

* Corresponding author.
E-mail address: piwczyn@umk.pl (M. Piwczyński).

https://doi.org/10.1016/j.ympev.2023.107913
Received 20 February 2023; Received in revised form 25 August 2023; Accepted 28 August 2023
Available online 1 September 2023

species, accounting for nearly half of the species diversity in the subfamily Gerbillinae (Musser and Carleton, 2005; Ndiaye et al., 2016a). The infra-generic taxonomy of the genus is highly debated. Traditionally, the genus was divided into three subgenera—*Gerbillus*, *Dipodillus*, and *Hendecapleura*—based on morpho-anatomical features (Lataste, 1882, 1881). However, starting from the beginning of the 21st century, *Gerbillus* and *Dipodillus* have been considered as distinct genera, primarily differentiated by their hairy or naked feet (Musser and Carleton, 2005; Pavlinov, 2008). Furthermore, the genus *Monodia* was established to accommodate the morphological distinctiveness of *G. mauritaniae*—a species described based on a single individual (Petter, 1975). The systematics of *Gerbillus* s.l. has also benefited from the inclusion of chromosomal information, as this group exhibits extensive interspecific karyotypic variability (Lay, 1983). Based on this, at least two evolutionary groups can be distinguished, which align with the *Gerbillus* and *Dipodillus* + *Hendecapleura* groups recognized based on morpho-anatomical data (Aniskin et al., 2006).

The last two decades have seen the confrontation of phenotype-based taxonomic systems with molecular information. In Gerbillinae rodents, the majority of the studies have solely relied on the mitochondrial cytochrome *b* gene (cyt-b) for both phylogenetic inference and species delimitation (Abiadh et al., 2010; Bryja et al., 2022; Ndiaye et al., 2014, 2012). Interestingly, these studies corroborated the subdivision of the genus *Gerbillus* into three subgenera: *Gerbillus, Dipodillus* and *Hendecapleura*. Furthermore, the subgenus *Hendecapleura* was placed as a sister group to the other subgenera making *Dipodillus* firmly nested within the genus *Gerbillus.* The most comprehensive study to date, which utilized cyt-b and one nuclear gene (inter-photoreceptor retinoid-binding protein or IRBP), also proposed the inclusion of *G. nancillus* in a new subgenus named *Monodia* (Ndiaye et al., 2016a). The choice of this name followed the work of Tranier and Julien-Laferriere (1990), who suggested the possible synonymy of *G. mauritaniae* (Heim de Balsac, 1943) and *G. nancillus*. Despite yielding coherent phylogenetic results, previous studies are limited by the use of only a single or, at most, two molecular markers, which may not provide sufficient phylogenetic signal to resolve certain parts of the tree. Additionally, there is a high risk that the resulting tree might not accurately represent the species tree as it is derived from a small sample of nuclear and mitochondrial genomes. Moreover, mitochondrial-based phylogeny might be further confounded by factors such as hybridization, mitochondrial introgression, heteroplasmy and presence of nuclear mitochondrial pseudogenes (Numts) (Bensasson et al., 2001; Bonnet et al., 2017; Dubey et al., 2009; Richly and Leister, 2004).

Accurate species delimitation is of critical importance for identifying endangered species, revealing cryptic variation and understanding patterns of diversification. In the case of *Gerbillus*, species have traditionally been identified and described using morpho-anatomical traits. However, morphological characters, such as the presence or absence of hairs on feet (Lay, 1983) or skull characters and associated muscle attachments (Alhajeri, 2021), can undergo adaptive and convergent evolution in response to similar selective pressure in similar habitats. Relying solely on phenotypic data may therefore underestimate the number of species and fail to identify cryptic variation. For example, in an attempt to delineate several *Gerbillus* species, Ndiaye et al. (2012) used 11 skull traits but found no clear boundaries among studied taxa. Consequently, the utilization of molecular data, specifically the cyt-b gene, was necessary to provide an independent source of information regarding species boundaries. Generally, molecular analyses indicate that all traditionally recognized species exhibit relatively high genetic divergence, forming monophyletic groups on the tree (Abiadh et al., 2010; Ndiaye et al., 2016a, 2012). However, these analyses assume the presence of barcode gaps, which can be misleading when species divergence is recent or when population size is large (DeSalle et al., 2005; Moritz and Cicero, 2004; Will and Rubinoff, 2004). To address these potential problems, the results of analyses based on single markers must be complemented with results from a multilocus coalescence-based delimitation method, which has recently become a standard tool for inferring species boundaries (Rannala and Yang, 2020).

The molecular data collected on *Gerbillus* have also facilitated the tentative reconstruction of the evolutionary history of the genus (Ndiaye et al., 2016a). Molecular dating indicates that the origin of the genus coincided with the establishment of arid conditions in the Sahara Desert, approximately 7 to 5 Mya (Schuster et al., 2006; Swezey, 2009). The divergence leading to the main lineages within the genus occurred during the transition between the Pliocene and the Pleistocene. Interestingly, some gerbil fossils from this time period, discovered in various locations in East and North Africa, have been suggested as direct ancestors of extant species (Winkler et al., 2010). The emergence of modern species was largely limited to the Pleistocene, except for *G. campestris,* which likely originated in the late Pliocene, approximately 3 Mya. The speciation events may have been influenced here by climatic oscillations during the Pleistocene, with major aridification episodes dated roughly at 2.8 Mya, 1.8 Mya and 1 Mya (deMenocal, 2004, 1995). One major limitation of these molecular clock analyses is the reliance on only a few markers, as is the case with phylogeny reconstructions and species delimitations. While uncertainty is an inherent property of Bayesian divergence time estimation, it can be reduced by exapanding the size of the dataset (dos Reis and Yang, 2013). Therefore, it is expected that a reanalysis of divergence timing in *Gerbillus,* utilizing a large number of markers from the genome, will enhance the precision of the evolutionary history reconstruction.

Here, we test recent hypotheses concerning the scenario of evolution of *Gerbillus* using restriction site-associated DNA sequencing (RADseq). This method has proven particularly useful for inferring shallow phylogenies for taxa that are a few million years old (e.g., Eaton and Ree, 2013; Tripp et al., 2017; Wagner et al., 2013). However, its utility has also been demonstrated for reconstructing phylogenies of older groups (e.g., Eaton et al., 2017; Grzywacz et al., 2021; Hipp et al., 2014; Piwczyński et al., 2021). We focus here on two main issues. Firstly, we reconstruct the phylogeny of the genus using genomic data to test the relationships among the four subgenera described based on morphology, karyology and classical Sanger data. Secondly, we examine species boundaries for several morphologically similar species, where doubts persist as to whether they are true species or simply reflect divergences among allopatric populations. The first taxon of interest is a putative new species with an uncertain geographic and population status found in and around the Massa River valley area (Morocco), provisionally named *Gerbillus* sp1 (Ndiaye et al., 2016a). It was hypothesized, based on molecular data, that it is a true species, closely related to *G. tarabuli* and *G. occiduus* (Ndiaye et al., 2016a, 2012). The second pair of taxa are from the subgenus *Hendecapleura—G. nanus* and *G. amoenus*—proposed to be vicariant species based on molecular data, with the former distributed from Pakistan to the Arabian Peninsula and the latter only in North Africa (Ndiaye et al., 2016b, 2013). The third delimitation test targets *G. campestris* and *G. rupicola*, two morphologically similar taxa differing in chromosome number ($2n = 56$ and $2n = 52$ respectively; Granjon et al., 2002). Given their high genetic similarity based on cyt-b and IRBP sequence data, Ndiaye et al. (2016a) put forward two hypotheses regarding their respective status. On the one hand, 'rupicola' may represent a chromosomally differentiated population of polymorphic *G. campestris*, a species where chromosomal variation has already been documented. On the other hand, 'rupicola' may be confirmed as a separate species recently originated from *G. campestris*.

The last aim of this study is to put phylogenomic results into historical context. The dating analysis is performed to confirm, with higher precision than in previous research, the timing of origin and Plio-Pleistocene diversification of the genus *Gerbillus*.

2. Materials and methods

2.1. Species sampling

Eleven *Gerbillus* species from North Africa and the Middle East, including one putative new species, represented by 37 specimens and an outgroup, *Psammomys obesus* (3 specimens), were collected during field expeditions over the last three decades (Barros et al., 2018; Boratyński et al., 2017; Nokelainen et al., 2020; Table 1). Tissue samples (pieces of ears or liver) were collected from individuals captured either with live-traps (Extra-Large Sherman Kangaroo Rat and locally-made wire-mesh live traps) or with handheld nets, during the night. Initial species identification was based on morphological features such as hind foot length, body mass, body and tail lengths, ear size, presence of hair on foot soles, body coloration, and proportion of body parts. In some cases, it was further confirmed by molecular barcoding (see below). All samples were georeferenced using a global positioning system. Animal handling and sampling followed relevant guidelines and regulations, and all study protocols were approved by local authorities (Haut Commissariat aux Eaux et Forêts et à la Lutte Contre la Désertification of Morocco, decisions 20/2013, 41/2014, 42/2014; Ministère de l'Environnement et du Développement Durable of Mauritania, decision 227/08.11.2012).

2.2. Barcoding

To validate morphology-based identification of *Gerbillus* specimens used in this study, we attempted to obtain the mitochondrial cyt-b gene, which serves as a standard barcode sequence in mammals (Bradley and Baker, 2001). Total genomic DNA was extracted from ca. 25 mg of tissue using the DNeasy Blood & Tissue Kit (Qiagen, Valencia, California, U.S.A.) following the manufacturer's instructions. The ≈1140 bp fragment of the cyt-b gene was PCR-amplified using primers L14723 and H15915 from Abiadh et al. (2010). Each 25 μL polymerase chain reaction (PCR) contained 1 × PCR buffer, 0.2 mM of deoxynucleoside triphosphates (dNTPs), 0.2 μM of each primer, 2.5–3.5 mM of MgCl2, 0.2 U of Taq DNA polymerase (ThermoScientific), and 1 μL of DNA template. The PCR protocol included an initial denaturation at 94 °C for 4 min, followed by 35 cycles with 40 s of denaturation at 94 °C, 45 s of annealing at 50 °C, and 40 s of extension at 72 °C. The final extension at 72 °C lasted for 5 min. For some troublesome samples, a gradient PCR was used to determine the optimal annealing temperature (48.0 °C, 48.9 °C, 49.8 °C, 51.1 °C, 52.4 °C, 54.9 °C).

After amplification, each PCR product was electrophoresed using 1% agarose gel and stained with GelRed (Biotium) to assess the success of PCR reactions. A PCR product was considered appropriate for sequencing if no obvious polymorphism (multiple bands from a single PCR product) was observed. The PCR product was then purified using Ampure (1:1.8) and sequenced using fluorescent Big Dye terminators (Applied Biosystems, Foster City, CA, USA). The sequencing products were precipitated using 0.3 M sodium acetate, washed with 80% ethanol, dried, and resolved using an automated DNA sequencer at the Laboratory of Molecular Biology Techniques, Adam Mickiewicz University (Poznań, Poland). The sequences were assembled and edited using SeqMan II v. 4.0 (DNASTAR, Madison, WI, USA). All newly obtained sequences were deposited in GenBank (Table 1).

Sequences obtained in this study were combined with cyt-b sequences for species of *Gerbillus* retrieved from GenBank. The sequences were then aligned using MUSCLE v.3.8.31 (Edgar, 2004) through the graphical interface in Seaview v.4.6.1 (Gouy et al., 2010). To examine whether cyt-b conspecific sequences clustered together, we performed a phylogenetic analysis using maximum likelihood method in RAxML v.8.2.12. The analysis included 100 rapid bootstrap repetitions under GTR + Γ substitution model, which is the only model available in the software (Stamatakis, 2014).

Table 1
Specimens used for generating RAD-seq data, along with corresponding voucher information and GenBank reference numbers for newly obtained cyt-b markers.

Taxon	Voucher information
Gerbillus amoenus (de Winton, 1902)	Mauritania, 21°01′04.8″ N 11°55′29.6″ W, 31 October 2011, leg. *Boratyński et al.* ZBSC0229, GenBank no: OQ230799
Gerbillus amoenus (de Winton, 1902)	Morocco, 23°34′48.0″ N 15°13′57.2″ W, 1 November 2012, leg. *Boratyński eBt al.* ZBSC0299, GenBank no: OQ230800
Gerbillus campestris Loche, 1867	Mauritania, 21°31′10.0″ N 12°51′10.2″ W, 25 October 2011, leg. *Boratyński et al.* ZBSC0213, GenBank no: OQ230801
Gerbillus campestris Loche, 1867	Morocco, 22°36′39.9″N 14°28′15.3″W, 5 January 2014, leg. *Boratyński et al.* ZBSC0416
Gerbillus campestris Loche, 1867	Mauritania, 18°22′02.9″ N 9°02′54.7″ W, 2 February 2014, leg. *Boratyński et al.* ZBSC0507
Gerbillus campestris Loche, 1867	Mauritania, 18°09′01.8″ N 12°03′56.6″ W, 8 February 2014, leg. *Boratyński et al.* ZBSC0567, GenBank no: OQ230802
Gerbillus gerbillus (Olivier, 1801)	Mauritania, 18°29′22.5″ N 14°38′37.6″ W, 27 November 2014, leg. *Brito et al.* 11405, GenBank no: OQ230803
Gerbillus gerbillus (Olivier, 1801)	Mauritania, 17°35′21.0″N 7°26′45.5″ W, 28 January 2014, leg. *Boratyński et al.* ZBSC0490, GenBank no: OQ230804
Gerbillus gerbillus (Olivier, 1801)	Mauritania, 18°23′01.4″ N 8°31′18.0″ W, 31 January 2014, leg. *Boratyński et al.* ZBSC0499
Gerbillus gerbillus (Olivier, 1801)	Mauritania, 20°43′49.5″ N 16°01′29.4″ W, 10 February 2014, leg. *Boratyński et al.* ZBSC0572
Gerbillus henleyi de Winton, 1903	Israel, 30°36′21.7″ N 34°50′44.6″ E, unknown date, leg. *Shenbrot et al.* H2
Gerbillus henleyi de Winton, 1903	Mali, 15°55′00.0″ N 2°28′00.0″ E, 21 February 2003, leg. *Papillon et al.* MEN3 (M4947)
Gerbillus henleyi de Winton, 1903	Mauritania, 16°45′46.6″ N 11°13′19.1″ W, 22 November 2012, leg. *Boratyński et al.* ZBSC0369, GenBank no: OQ230805
Gerbillus nancillus Thomas & Hinton, 1923	Senegal, 15°51′20.5″ N 15°03′47.2″ W, 5 August 2010, leg. *Bà et al.* KB7364 (S-KB7364)
Gerbillus nancillus Thomas & Hinton, 1923	Mali, 15°11′00.5″ N 9°31′39.7″ W, 30 November 1999, leg. *Granjon et al.* M4067
Gerbillus nancillus Thomas & Hinton, 1923	Mauritania, 17°25′22.2″ N 13°26′06.6″ W, 17 November 2010, leg. *Boratyński et al.* ZBSC0058
Gerbillus nanus Blanford, 1875	Jordan, 29°58′00.1″ N 35°04′60.0″ E, 10 May 1995, leg. *Benda et al.* 759 (NMP48239)
Gerbillus nanus Blanford, 1875	Israel, 30°43′30.4″ N 35°16′20.8″ E, unknown date, leg. *Shenbrot et al.* N2
Gerbillus nanus Blanford, 1875	Israel, 30°43′30.4″ N 35°16′20.8″ E, unknown date, leg. *Shenbrot et al.* N3, GenBank no: OQ230806
Gerbillus occiduus Lay, 1975	Morocco, 23°55′05.0″ N 15°45′48.0″ W, 1 December 2008, leg. *Granjon et al.* LG130
Gerbillus occiduus Lay, 1975	Morocco, 23°53′06.2″ N 15°49′48.3″ W, 1 December 2008, leg. *Granjon et al.* LG135
Gerbillus occiduus Lay, 1975	Morocco, 23°53′06.2″ N 15°49′48.3″ W, 1 December 2008, leg. *Granjon et al.* LG140, GenBank no: OQ230807
Gerbillus pyramidum (É. Geoffroy Saint-Hilaire, 1803)	Mali, 19°19′60.0″ N 0°14′30.0″ W, 19 November 2005, leg. *Granjon et al.* M−INA20
Gerbillus pyramidum (É. Geoffroy Saint-Hilaire, 1803)	Mali, 19°01′10.3″ N 1°45′42.8″ E, 7 February 2004, leg. *Granjon et al.* M5354
Gerbillus pyramidum (É. Geoffroy Saint-Hilaire, 1803)	Mali, 20°11′39.2″ N 0°58′20.1″ E, 14 November 2005, leg. *Granjon et al.* M5963
Gerbillus pyramidum (É. Geoffroy Saint-Hilaire, 1803)	Mauritania, 19°48′31.7″N 14°17′18.5″W, 20 November 2010, leg. *Boratyński et al.* ZBSC0066, GenBank no: OQ230808
Gerbillus pyramidum (É. Geoffroy Saint-Hilaire, 1803)	Mauritania, 21°01′04.8″ N 11°55′29.6″ W, 1 November 2011, leg. *Boratyński et al.* ZBSC0234
Gerbillus pyramidum (É. Geoffroy Saint-Hilaire, 1803)	Mauritania, 21°01′04.8″ N 11°55′29.6″ W, 1 November 2011, leg. *Boratyński et al.* ZBSC0235
Gerbillus rupicola (Granjon, Aniskin, Volobouev & Sicard, 2002)	Mali, 14°12′51.4″ N 3°54′09.3″ W, 7 March 2008, leg. *Cosson et al.* M4713, GenBank no: OQ230809

(*continued on next page*)

Table 1 (*continued*)

Taxon	Voucher information
Gerbillus rupicola (Granjon, Aniskin, Volobouev & Sicard, 2002)	Mali, 14°04′27.4″ N 4°08′23.9″ W, 19 January 2003, leg. *Sicard et al.* M4936
Gerbillus sp1	Morocco, 30°04′00.0″ N 9°39′30.0″ W, 23 November 2008, leg. *Granjon et al.* LG77
Gerbillus sp1	Morocco, 29°49′17.3″ N 9°49′23.5″ W, 24 November 2005, leg. *Granjon et al.* LG91, GenBank no: OQ230810
Gerbillus sp1	Morocco, 29°49′17.3″ N 9°49′23.5″ W, 24 November 2005, leg. *Granjon et al.* LG96
Gerbillus tarabuli (Thomas, 1902)	Morocco, 27°09′12.5″ N 10°50′50.0″ W, 29 October 2012, leg. *Boratyński et al.* ZBSC0288, GenBank no: OQ230811
Gerbillus tarabuli (Thomas, 1902)	Mauritania, 16°29′02.9″ N 9°17′22.5″ W, 26 January 2014, leg. *Boratyński et al.* ZBSC0486,
Gerbillus tarabuli (Thomas, 1902)	Morocco, 32°15′14.3″ N 2°11′16.4″ W, 7 May 2014, leg. *Boratyński et al.* ZBSC0592
Gerbillus tarabuli (Thomas, 1902)	Morocco, 32°15′26.5″ N 2°14′26.6″ W, 7 May 2014, leg. *Boratyński et al.* ZBSC0593, GenBank no: OQ230812
Psammomys obesus Cretzschmar, 1828	Morocco, 27°54′03.5″ N 11°36′18.4″ W, 15 February 2015, leg. *Boratyński et al.* ZBSC0655
Psammomys obesus Cretzschmar, 1828	Morocco, 26°40′27.8″ N 12°56′52.3″ W, 21 February 2015, leg. *Boratyński et al.* ZBSC0671
Psammomys obesus Cretzschmar, 1828	Morocco, 28°00′02.7″N 12°13′09.6″W, 13 August 2015, leg. *Boratyński et al.* ZBSC0711, GenBank no: OQ230813

2.3. RAD-seq library preparation and data processing

A reduced complexity library was prepared based on the traditional RAD-seq protocol described by Ali et al. (2016) with the following modifications (see also Grzywacz et al., 2021; Piwczyński et al., 2021):

i) For each individual, two samples of 75 ng of DNA each were separately digested using SbfI-HF restriction enzyme to avoid unexpected reaction failure.

ii) P1 adapter-ligated fragments were sheared for 60 s to a peak target of 300 bp using an ultrasonicator Covaris M220 (Covaris, Inc.).

iii) Pippin Prep (Sage Science, Beverly, MA, U.S.A.) was used to select fragments between 250 and 350 bp with prior library cleaning using AMPure XP (Beckman Coulter, Carlsbad, CA, U.S.A.; 1 × ratio of bead to sample volume).

iv) Eight independent PCRs (15 cycles) were carried out and subsequently pooled to reduce the effect of uneven fragment amplification.

v) Pooled PCR products were purified twice with AMPure XP (1 × ratio of bead to sample volume) to completely remove the remaining primers.

The final library quantification was performed using a Qubit 3.0 fluorometer and 2100 Bioanalyzer with the High Sensitivity DNA Analysis Kit (Agilent Technologies). Commercial paired-end sequencing of the multiplexed library was conducted on an IlluminaHiSeq 2500 instrument by the Macrogen company.

Before running the assembly software, we filtered the low quality reads with Trimmomatic v.0.39 (Bolger et al., 2014) using the following options: *TRAILING:3 SLIDINGWINDOW:4:20 MINLEN:50*. We deliberately did not set the *LEADING* option to preserve the barcode part of the reads before applying *SLIDINGWINDOW* option. For the *de novo* and reference assembly, we utilized only R1 data with the ipyrad v.0.9.61 pipeline (Eaton, 2014). It is worth noting that, as confirmed by ipyrad developers (see discussion on Gitter, 25 September 2020, https://gitter.im/dereneaton/ipyrad), and based on our own experience (Grzywacz et al., 2021; Piwczyński et al., 2021), excluding R2 reads typically does not significantly impact the final result.

In the first step, we demultiplexed the data twice allowing none or one mismatch in the barcode sequences. Taking into account the sensitivity of the *de novo* assembly to changes in clustering threshold (CT) parameter (i.e., the minimum percentage of sequences similarity below which two reads are considered to have come from different loci; Fernández-Mazuecos et al., 2018; Shafer et al., 2017; Takahashi et al., 2014), we tested, using two demultiplexed datasets, a wide range of CT from 0.85 to 0.99 incremented by 0.01 resulting in a total of 15 analyses per dataset. The remaining parameters were set as follows: *min_samples_locus* = 4 and *max_Indels_locus* = 8. Given the high proportion of repetitive elements in rodents' genome (up to 45%, Lu et al., 2020) and the extensive karyotype reorganization among *Gerbillus* species (Aniskin et al., 2006), we adopted a conservative approach by reducing *max_SNPs_locus* parameter to 0.1 (compared to the default value 0.2) to minimize potential alignment issues with repetitive regions.

We employed two criteria to select the best CT: the average bootstrap support and the number of SNPs. The average bootstrap support was estimated based on a phylogenetic tree reconstructed using maximum likelihood (ML) approach implemented in RAxML v.8.2.12 with 100 rapid bootstrap repetitions (Stamatakis, 2014).

To assess the impact of missing data on the phylogenetic tree topology, we explored three additional values of *min_samples_locus* = 12, 20 and 32 corresponding to a maximum of 70%, 50% and 20% of missing data in the alignments, respectively. For each *min_samples_locus* value, we tested the entire range of CT values, i.e., 0.85–0.99, and selected the best value based on the previously described criteria. The remaining parameters were left unchanged, as explained earlier. All analyses were conducted on two demultiplexed datasets.

In the final set of analyses, we used two demultiplexed datasets for reference assembly, employing the genome of *Psammomys obesus* (GenBank: GCA_907164565). We generated assemblies using two values for *max_SNPs_locus* = 0.1 and 0.6. The relaxation of this parameter might introduce confounding phylogenetic signal by treating paralogs as orthologs or it may enhance the signal by providing more variable orthologous loci. The remaining parameters were kept unchanged, as previously mentioned.

2.4. Phylogenetic analyses of RAD-seq data

The data generated by ipyrad with the highest average bootstrap support value and the highest number of SNPs from *de novo* analyses with *min_samples_locus* = 4, 12, 20 and 32 as well as the data from the reference-based assembly (see above) were subjected to analyses using both maximum likelihood (ML) and Bayesian (BI) approaches as implemented in RAxML v.8.2.12 (Stamatakis, 2014) and ExaBayes v.1.5.1 (Aberer et al., 2014), respectively. For all ML analyses, 100 searches of tree space were performed, each starting from distinct randomized maximum parsimony tree. Branch support (BS) was evaluated based on 1 000 standard non-parametric bootstrap replicates and then summarized on the best ML tree. For Bayesian analyses, six simultaneous runs were completed each using a single Monte Carlo Markov chain. Each chain was run for 1 million generations, with samples collected every 1 000 generations and the first 25% of samples discarded as burn-in (a total of 6 [runs] × 1[chains] × 750[samples – burn-in] = 4 500 samples across all runs). All priors remained unmodified and in their default state. Effective sampling of the priors and MCMC convergence was established by ensuring effective sample size (ESS) and potential scale reduction factors (PRSF) were > 200 and ~ 1, respectively, using Tracer v.1.7.1 (Rambaut et al., 2018). ML analyses were conducted via CIPRES science gateway (Miller et al., 2010), while the BI analyses were completed using the University of Technology, Sydney eResearch High Performance Computing Cluster.

The main assumption of ML and BI methods applied here to concatenated datasets is that all sites across the genome share the same history. However, this assumption carries a risk of reconstructing a highly supported, yet incorrect species tree (Mendes and Hahn, 2018). To address this concern and consider potential gene tree heterogeneity due to incomplete lineage sorting, we utilized the SVDquartets method (Chifman and Kubatko, 2014), implemented in the most recent version of PAUP* v.4.0a (Swofford, 2003) for sequence data under the

multispecies coalescent model. For the analyses, SNP files automatically generated by ipyrad were used, containing one randomly selected SNP per locus. We evaluated all possible quartets (exhaustive quartet sampling option) and treated ambiguities as missing data. The QFM algorithm was then applied to assemble quartets into a species tree (Reaz et al., 2014). Confidence in the groupings on trees was measured using standard non-parametric bootstrap procedure with 1 000 repetitions.

Commonly used branch support methods for genomic data, such as non-parametric bootstrap support or Bayesian posterior probability, may lead to an overestimation of support when the underlying models of sequence evolution are overly simple (Pease et al., 2018). This issue becomes especially relevant in concatenation analyses where data, comprising sequences from various parts of the genome, are treated as a single large locus. Even in the case of SVDquartets analysis, bootstrap support values may be unreliable when the underlying model of sequence evolution is complex and branch lengths (in coalescence units) are short (see fig. 5 in Chifman and Kubatko, 2014). To address these challenges and supplement traditional methods, we employed the quartet sampling (QS) method (Pease et al., 2018) to evaluate branch support for trees resulting from concatenation analyses. QS takes both a phylogenetic tree and an alignment as input, assessing three scores: quartet concordance (QC), quartet differential (QD), and quartet informativness (QI). Scores are estimated for each internal branch by the relative frequency of the three possible quartets (for a detailed technical and mathematical description refer to Appendix S1 in Pease et al., 2018). Briefly, QC reports how often the sampled quartet topology concurs with the input tree. This score ranges from 1 (when all sampled quartets are concordant) to −1 (when all are discordant), with 0 indicating an equal frequency of the three quartet topologies. QD relies on the stochastic process model of incomplete lineage sorting where the probabilities of two discordant quartet topologies are expected to be similar. If there is, for example, introgression, one of the discordant topologies will be enriched in frequency relative to the other. Values for QD fall in the range [0,1], with maximal value indicating an identical proportion of discordant trees and minimal value suggesting that only one of the two discordant topologies was sampled. QI quantifies the phylogenetic signal and ranges from [0,1]. A quartet is considered informative when its likelihood estimated using the ML method (in our case, RAxML) is higher than some cut-off value from the second-best quartet topology. In the extreme case, when topologies have indistinguishable likelihood values, QI equals 0. All three scores should be interpreted together to distinguish different sources of uncertainty, aided, for example, by interpretational guidelines provided by Pease et al. (2018) (see Table 1 therein). To estimate all three scores, we used python software quartetsampling v.1.3.1 (https://github.com/FePhyFoFum/quartetsampling; accessed Ferbruary 2022) with the following parameters: *–reps* 250 (number of replicates per internal branch), *–lnlike* 2 (likelihood cut-off value) and *–engine* raxml (program to estimate tree likelihood).

2.5. Species delimitation

We used Bayes Factor Delimitation (BFD) in SNAPP v1.5.2 (Grummer et al., 2014; Leaché et al., 2014), a module within the BEAST v2.6.6 software (Bouckaert et al., 2019), to determine the most plausible assignment of individuals to species using a multispecies coalescent model with no gene flow. BFD requires a priori species models, which are competing assignment of individuals to species. These models are further analyzed separately using SNAPP and ranked by comparing Bayes factor (BF). In our analysis, we considered four species delimitation models. In the first model, which we considered as a full model, individuals were assigned to species based on the original morphological determination, including the putative new species. In the subsequent models (two to four), we merged individuals representing the following species: *G. occiduus* and *Gerbillus* sp1 in the second model, *G. campestris* and *G. rupicola* in the third model, and *G. amoenus* and *G. nanus* in the fourth model.

Since consistent phylogenetic trees were obtained from data generated under various parameters in ipyrad (see Results section), we simply opted for the one with the lowest amount of missing data as a source of SNPs for SNAPP analyses. To ensure the independence of sampled SNPs, we converted the unlinked SNPs file from ipyrad (.usnps) to a three-state numeric (0, 1 and 2) nexus file suitable for SNAPP. For this purpose, we utilized the python script unlinked_snps_to_nexus.py, available at https://github.com/brunoasm/usnps_to_nexus (accessed 13 April 2022). After conversion, we faced a large number of loci (over 12 000), making it impractical to analyze them by the software in a reasonable time frame. Therefore, using custom R script, we sampled only those loci that did not have missing data. As a result, the final input file comprised 2 650 biallelic loci.

For each SNAPP analysis, mutation rates u and v were fixed at 1 and not sampled. We tested two gamma priors for speciation rate λ: $\Gamma(\alpha = 2.0, \beta = 200.0)$ and $\Gamma(\alpha = 2.0, \beta = 20.0)$, as well as two for effective population size for ancestral populations given in terms of population mutation rate θ: $\Gamma(\alpha = 1.0, \beta = 2\,500.0)$ and $\Gamma(\alpha = 1.0, \beta = 250.0)$. In the case of λ, the first prior assumes bifurcation every 0.0025 substitutions per site, resulting, on average, in 400 speciation events every substitution per site of tree length. In contrast, the second prior assumes bifurcation every 0.025 substitutions per site, resulting, on average, in only 40 speciation events. The 2.5% and 97.5% quantiles range from 48.4 to 1.11×10^3 in the first prior and from 4.84 to 111 in the second prior. For population mutation rate θ, the first prior assumes that for two randomly selected sequences within a population, one expects, on average, 4 heterozygous sites per 10 kbp, while in the second assumes 4 heterozygous sites per 1 kbp. The lower and upper quantiles here are 1.0×10^{-5}–1.5×10^{-3} and 1.0×10^{-4}–1.5×10^{-2}, respectively. We tested all combinations (4 sets of analyses) of λ and θ priors for our delimitation models. Each alternative model was analyzed using 50 steps of path sampling with 100 000 MCMC steps each and 10 000 pre-burn-in steps to estimate marginal likelihood. The first 25% of MCMC generations were discarded as burn-in. To identify the most likely species delimitation, we compared the estimated marginal likelihoods for competing models using $BF = 2 \times (MLE_1 - MLE_2)$. We consider the model decisively better supported by data when $BF > 10$.

All species delimitation analyses were completed using the computing cluster FUN-K at Biological and Chemical Research Center, University of Warsaw.

2.6. Molecular dating

Due to the lack of reliable fossils of *Gerbillus* suitable as calibration points in this study, we used three secondary calibration points obtained from a Bayesian analysis of 161 species representing all major lineages of murid rodents (Aghová et al., 2018) (Table 2). We selected this particular study because it overcomes the limitations of other phylogenetic analyses, such as Steppan and Schenk (2017). Notably, the phylogenetic tree was constructed using an alignment with 42.5% missing data, in contrast to 65% in Steppan and Schenk (2017). Additionally, the dating was performed using BEAST, enabling us to extract secondary calibration points, which was not feasible in Steppan and Schenk (2017) due to convergence issues in their BEAST analysis. The choice of fossil calibration points followed the rigorous recommendations of Parham et al. (2012), and all calibration points underwent cross-validation using the procedure of Near and Sanderson (2004). Most importantly, Aghová et al., 2018 used two fossils (*Abudhabia pakistanensis* and *Gerbilliscus* sp.) in their final analysis, both closely related to *Gerbillus*. This becomes particularly important as it allows us to expect relatively precise time estimates in proximity to *Gerbillus*.

The first calibration point was placed on the root node, representing the common ancestor of *Gerbillus* and *Psammomys*. As *Psammomys* was not included in the phylogenetic analysis of murid rodents, we relied on information from the node representing the common ancestor of

Table 2
Secondary calibration points used in BEAST analyses of the genus *Gerbillus*, based on a phylogeny of 161 species representing all the major lineages of murid rodents (Aghová et al., 2018). The mean (μ) and variance (σ^2) of user-specified log-normal calibration priors were estimated using *fit.perc* function in the R library rriskDistributions, based on quantiles obtained from the original analysis of Aghová et al. (2018). An adjustment was made to correct for bias between means of effective prior densities and user priors (see also Fig. A.1). The final dating analyses were performed using corrected user-specified calibration priors. All parameters are given on the real scale in million years.

Calibration points	Quantiles (%)			User-specified priors		Effective priors		Corrected user-specified priors		Effective prior after correction	
	2.5	50	97.5	μ	σ^2	μ	σ^2	μ	σ^2	μ	σ^2
Root	10.46	12.16	14.16	12.18	1.08	11.41	1.08	12.95	1.08	12.10	1.08
Genus *Gerbillus*	3.85	4.79	5.91	4.81	1.12	4.94	1.10	4.67	1.12	4.80	1.11
Monodia and *Dipodillus/Gerbillus*	2.89	3.66	4.56	3.67	1.13	3.91	1.11	3.43	1.13	3.71	1.11

Gerbillus and *Gerbilliscus/Desmodillus* (Aghová et al., 2018). Both splits, *Gerbillus/Gerbilliscus* and *Psammomys/Gerbillus,* occurred very close in time to each other, likely in the Upper Miocene (11.63–5.33 mya) (Steppan and Schenk 2017). The remaining two calibration points were set to the common ancestor of the genus *Gerbillus* and the common ancestor of the subgenera *Monodia* and *Dipodillus/Gerbillus,* respectively. The calibration point on *Gerbillus* corresponds to the earliest fossil of the genus, dated in the early Pliocene (ca. 4.1 Mya; Winkler et al., 2010). We assumed that all prior probabilities for the calibration points followed log-normal distributions. Since variance was not available from the published tree (Aghová et al., 2018), we used quantiles (2.5, 0.5 and 97.5) to fit the log-normal distribution and calculate first and second moments. This was done using *fit.perc* function in the R package rriskDistributions (https://cran.r-project.org/web/packages/rriskDistributions/index.html, accessed 17 June 2022).

Estimation of divergence times was performed using BEAST v.2.6.6 (Bouckaert et al., 2019) with a GTR + Γ substitution model and the uncorrelated lognormal (UCLN) model for substitution rates across lineages. We used the optimized relaxed clock algorithm, implemented in the ORC package for BEAST 2, which more efficiently explores relaxed clock parameters compared with previous setups (Douglas et al., 2021). The rate of cladogenesis was assumed to follow a Yule process for which we specified gamma priors with shape parameter set to 0.001 and scale set to 1 000. The priors for all other parameters were left at their default values. Two independent Markov chains were run, each for 10 000 000 generations and sampled every 1 000 generations.

Prior to data analysis, initial runs were performed without data to generate so-called effective priors on calibration points used by the software. The effective priors can differ significantly from the original user-specified calibration densities due to their dependence on the interaction among user priors, topology, and the birth–death process that defines the distribution of the ages of the non-calibrated nodes (Barba-Montoya et al., 2017; Inoue et al., 2010; Warnock et al., 2015). Based on these results, we compared effective priors with user-specified ones and made adjustments when necessary (Table 2, Fig. A.1).

The log files were examined using Tracer v.1.7.2 (Rambaut et al., 2018) to verify convergence and effective sample size. Tree files from all runs were combined using LogCombiner and summarized with TreeAnnotator. The first 25% of the saved trees from each run were discarded as burn-in, and the remaining 15 000 trees were summarized by choosing a maximum clade-credibility tree as a target tree with node heights estimated using mean values and a posterior probability threshold of 0.5 for the estimation of the 95% highest posterior density interval. The resulting tree was visualized using FigTree v.1.4.4 (https://tree.bio.ed.ac.uk/software/figtree/, accessed June 2022).

The supplementary material A contains all the command lines and scripts used to execute the analyses.

3. Results

3.1. Barcoding

Consistent with problems regularly encountered with traditional molecular procedure to obtain the cyt-b gene from *Gerbillus* (Bryja A. and Bryja J., personal communication), we successfully obtained only 14 sequences, representing 10 species of *Gerbillus* (including a putative new species) and one of *Psammomys obesus* (Table 1). We were unable to obtain the barcode for *G. nancillus*, as a pseudogene was amplified instead for each studied individual. After merging these sequences with those obtained from GenBank and filtering out short (<900 bp), poor quality and identical sequences, the final alignment contained 387 sequences representing 19 species of *Gerbillus* and two outgroups.

Five analyzed species—*G. nanus*, *G. henleyi*, *G. pyramidum*, *G. occiduus*, *G. gerbillus*—clustered together with their own conspecifics in well-supported clades (99–100%) (Fig. B.1). Two individuals representing *G. rupicola* were also placed in a clade containing only conspecifics, albeit with lower bootstrap support (76%). The putative new species (*Gerbillus* sp1) did not cluster with any of the analyzed species. Instead, it grouped with other undetermined specimens collected from the same area (Ndiaye et al., 2012), in a highly supported group (100%). Two individuals of *G. amoenus* were placed with other *G. amoenus* individuals from GenBank where they were intermingled with certain individuals of *G. nanus*. Because *G. nanus* formed a separate clade, the intermingled clade appears to be a historical artifact from the time when African sequences representing *G. amoenus* were described under the name of *G. nanus*. Currently, this name is restricted only to Asian individuals. Although *G. campestris* was placed in a strongly supported group (100%), it was not monophyletic, as *G. rupicola* individuals, from this study and from GenBank, were nested within it.

3.2. Assembly of RAD-seq data

A total of 433 411 452 reads (R1 + R2) were generated by Illumina sequencing of the RAD-seq library. The percentage of bases with quality scores Q20 and Q30 in the dataset was 95.85 and 92.18, respectively. After filtering out low quality reads and those potentially containing Illumina adapters with Trimmomatic, 191 544 678 paired reads were retained, along with 13 107 415 and 5 324 807 unpaired for R1 and R2, respectively. For further analyses, we used only R1 paired and unpaired reads. After demultiplexing R1 reads with ipyrad, the number of raw reads for studied specimens ranged from 247 542 for *Gerbillus nanus* 759 to 8 409 730 for *Psammomys obesus* ZBSC0655 when no mismatch in the barcode site was allowed. When one mismatch was permitted, the range changed from 257 798 to 8 583 974.

In the first set of ipyrad *de novo* analyses (*min_samples_locus* = 4, *max_Indels_locus* = 8 and *max_SNPs_locus* = 0.1), data generated under CT = 0.87 and 0.96 yielded the highest mean bootstrap support (avgBS = 95.35%) and the largest number of SNPs (SNP = 1 160 209), respectively, when no mismatches were allowed in the barcode sequence. With one mismatch allowed, CT = 0.95 and 0.96 provided the highest scores for our criteria (avgBS = 95.16% and SNP = 1 160 352; Table 3 and Fig. C.1). In the second set of analyses under different *min_samples_locus* values (12, 20 and 32), we obtained identical threshold values for both criteria only in the case of *min_samples_locus* = 20, which was 0.95 in both datasets (with and without mismatches in barcode sequences). For the remaining analyses, the best threshold

Table 3
Summary of *de novo* assembly of RAD-seq data under different values of *min_samples_locus* (msl) parameter, controlling the maximal number of missing data in a final alignment. Clustering threshold values ranging from 0.85 to 0.99 were tested for each *min_samples_locus* value. The clustering thresholds (CT-avgBS and CT-SNP) providing the highest average bootstrap values (avgBS) and SNP numbers (SNP) are shown. The empirical percent of missing data in final alignment for each analysis is also presented. Additionally, alongside average bootstrap values (avgBS), the median of bootstrap values (medianBS) is given. All analyses were carried out for the *max_barcode_mismatch* (mbm) parameter set to 0 or 1.

msl	CT-avgBS	avgBS/medianBS	% of missing data	CT-SNP	SNP	% of missing data
mbm = 0						
4	0.87	95.35/100	66.44%	0.96	1 160 209	72.72%
12	0.89	94.32/100	38.95%	0.96	569 015	48.38%
20	0.95	88.08/100	30.66%	0.95	311 197	30.66%
32	0.91	85.43/99	11.00%	0.94	97 606	11.47%
mbm = 1						
4	0.95	95.16/100	71.15%	0.96	1 160 352	72.75%
12	0.95	92.78/100	46.52%	0.96	567 396	48.28%
20	0.95	89.19/100	30.59%	0.95	312 382	30.59%
32	0.96	84.81/99	12.03%	0.94	99 182	11.42%

values differed between criteria, especially in the dataset with no mismatches allowed in barcode sequences (Table 3). The avgBS decreased along with the number of missing data comparably in both sets of analyses, reaching approximately 85% for *min_samples_locus* = 32. Similarly, number of SNPs dropped below 100 000 for analyses with the lowest number of missing data, irrespective of the applied settings to the barcode sequence. The median bootstrap values were in all cases close to or equal to 100%, indicating that a few poorly supported nodes strongly affect avgBS (Table 3).

The ipyrad reference analyses, under *max_SNPs_locus* = 0.1, resulted in 544 450 and 541 406 SNPs for datasets with and without mismatches in barcode sequences, respectively, and the resulted alignment matrices consisting of 67.60% and 67.42% missing data. Increasing the *max_SNPs_locus* parameter to 0.6 led to a much higher number of SNPs, 1 874 648 and 1 864 829, with less missing data in final alignments, 52.30% and 52.17%.

3.3. Phylogenetic analyses

When considering only interspecific relationships, the concatenation analyses with RAxML and ExaBayes consistently resulted in one topology, regardless of the number of mismatches allowed in barcode sequence, the amount of missing data in the final alignments, and the type of assembly (*de novo* or reference-based) (Fig. 1). In the case of RAxML, all branches except for two were strongly supported (BS = 96–100%), while Exabayes assigned maximal posterior probability values (PP = 1.0) to all branches. The two exceptional branches in ML analyses showed varying bootstrap support values, ranging from 1% to 100% for the first branch and from 80% to 100% for the second branch (Fig. 1).

In the majority of cases (30/40 analyses), SVDquartets produced identical topology to the concatenation analyses, although some branches exhibited more variable bootstrap support values (Fig. 2). However, in ten analyses, five alternative topologies were observed (Fig. D.1). These alternative topologies primarily involved changes around the branches '1' or '2' (Fig. 2). Notably, four topologies placed *Gerbillus nanus* 759 outside its conspecifics, and one topology differed in the placement of the *G. nancillus* clade (Fig. D.1).

The dominant topology grouped all specimens identified as belonging to one morpho-species into clades. These clades formed groups which correspond to the four traditional subgenera recognized in *Gerbillus*: *Dipodillus*, *Gerbillus*, *Hendecapleura* and *Monodia*. The representatives of the subgenus *Hendecapleura* (*G. nanus, G. amoenus* and *G. henleyi*) formed a sister group to all the remaining congeners. The unstable branch '2' was located within this clade. The next split led to *G. nancillus*, representing here the subgenus *Monodia,* and the two remaining subgenera—*Dipodillus* and *Gerbillus*—which formed sister groups. The uncertain branch '1' connected the divergence point of these latter taxa. In the subgenus *Gerbillus*, the putative new species was clustered as a sister species to *G. occiduus*.

The QS scores provided strong support for focal branches and demonstrated a strong phylogenetic signal in all branches, except for two (Table 4 and Fig. E.1). These two branches corresponded to the ones with varying bootstrap support in ML analyses. For the first branch (referred to as branch '1' in Figs. 1–2), the high QI values, ranging from 0.90 to 1.00, indicated a sufficient amount of information for all sampled quartets. However, despite the dependency of QC and QD scores on analyzed datasets, they showed some degree of support or a lack of support for branch '1' (QC = -0.06–0.83) and strong to moderate support for alternative evolutionary histories (QD = 0.00–0.59) (Table 4). The interpretation of QS scores for the second branch (referred to as branch '2' in Figs. 1–2) was more complex. The QI score ranged from 0.72 to 1.00, indicating moderate information for sampled quartets in some cases. Furthermore, QC and QD scores showed mutually exclusive interpretations, with varying degrees of support for the focal and alternative branches (Table 4). For example, in some datasets, QS scores provided full support for branch '2' (QC = 1.0, QD = –, QI = 1.00), while in others, strong support was observed for alternative evolutionary histories (e.g., QC = 0.28, QD = 0.00, QI = 0.94) Likely, the low number of reads acquired for *G. nanus* 759 is responsible for these discordant results for branch '2', leading to the differing interpretations.

3.4. Species delimitation

Species delimitation analyses using the BFD method strongly supported a full species model, regardless of the priors set for speciation rate λ and population mutation rate θ (Table 5). The second-best supported scenarios were either models that combined *Gerbillus amoenus* and *G. nanus* [priors $\lambda = \Gamma(\alpha = 2.0, \beta = 200.0/20.0)$ and $\theta = \Gamma(\alpha = 1.0, \beta = 2500.0)$] or *G. occiduus* and *Gerbillus* sp1 [priors $\lambda = \Gamma(\alpha = 2.0, \beta = 200.0/20.0)$ and $\theta = \Gamma(\alpha = 1.0, \beta = 250.0)$] as one species. However, both scenarios received much lower support, with BF > 100 or BF > 200, respectively, in comparison with the full model (Table 5).

3.5. Molecular dating

Bayesian analyses with BEAST resulted in a tree having a congruent topology to that obtained in ML and BI analyses (Fig. 3). The estimated time for the root was 11.78 Ma, with a 95% highest posterior density interval (HPD) of 9.98–13.58 Ma. The divergence of *Gerbillus* occurred at 4.66 Ma (95% HPD: 3.76–5.60 Ma). The sole representative of subgenus *Monodia*, *G. nancillus*, separated from subgenera *Dipodillus* and *Gerbillus* around 3.6 Ma (95% HPD: 2.87–4.36 Ma). The divergence of the two latter groups took place around 3.17 Ma (95% HPD: 2.20–4.09 Ma). The putative new species, *Gerbillus* sp1, separated from *G. occiduus*, its closest relative in this study, about 1.09 Ma (95% HPD: 0.32–1.96 Ma). Generally, sister species diverged approximately 1–2 Ma.

msl	BS-1	BS-2
4	96/8:44/23	100/100:100/100
12	4/6:2/1	99/99:99/97
20	2/5*	99/98*
32	100/97:100/100	80/93:98/98
Reference	78/72:100/100	97/99:96/99

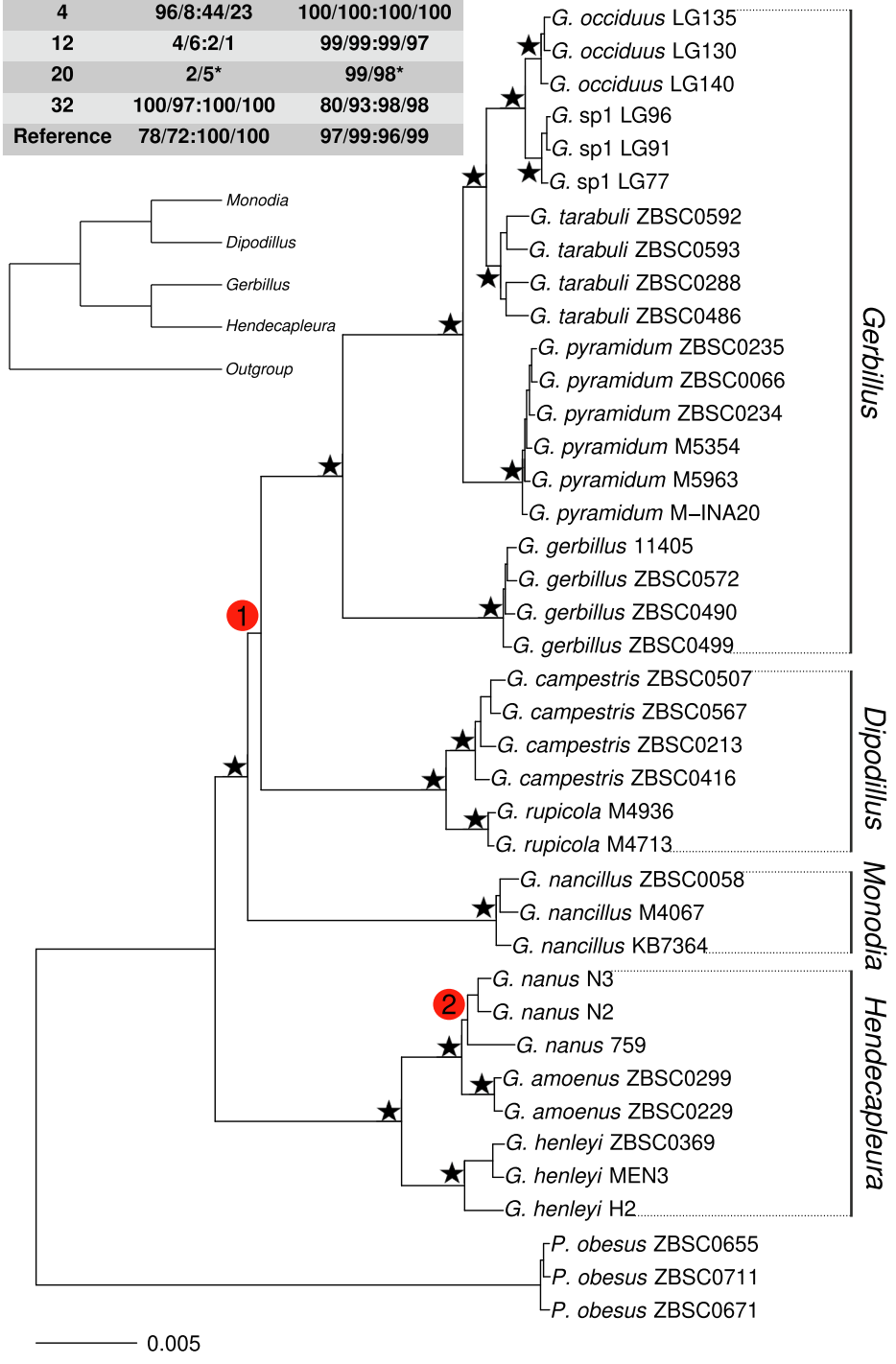

Fig. 1. An exemplary maximum likelihood tree inferred from the analysis of sequences generated from RAD-seq data for *Gerbillus*. This topology, omitting intraspecific variation, was obtained for all ML and BI analyses of each dataset produced by ipyrad under various parameter settings. All branches, except two (branch '1' and '2' in the figure), were strongly supported (96–100%) by ML analyses and marked by stars. The table presents the variation in bootstrap values (BS-1 and BS-2) for the two marked branches in reference and *de novo* analyses obtained under different values of *min_samples_locus* (msl). Bootstrap values are shown for the best datasets, in terms of average bootstrap support and number of SNPs (separated by colon), generated in ipyrad under a range of clustering threshold values allowing *max_barcode_mismatch* to be 0 or 1 (separated by slash). For msl = 20 (marked by stars), the dataset obtained for the same clustering threshold value was considered the best by both bootstrap and SNP criteria. The major clades corresponding to the traditionally recognized subgenera are bracketed. The representatives of *Psammomys obesus* were treated as outgroup. The alternative history supported by QC score for branch '1' is also shown. For all BI analyses (not shown), the posterior probability value for each branch was equal 1.0.

4. Discussion

Recent attempts to study the diversification of *Gerbillus* have focused on three main themes: resolving the phylogeny, tracing the evolutionary history of the genus by correlating molecular dating with climatic/environmental oscillations, and assessing species diversity using molecular data (e.g., Abiadh et al., 2010; Ndiaye et al., 2016a, 2016b). However, all studies were limited by the number of markers used, with cyt-b being the primary source of phylogenetic information. Surprisingly, despite this limitation, the majority of conclusions from this research were corroborated here by the analysis of genomic data. We center our discussion around these three themes, suggesting the possible biological causes of the congruence among various data sets.

4.1. Concatenation and coalescence methods confirm the division of the genus Gerbillus into four subgenera but show alternative evolutionary histories for inter-subgeneric relationships

Previous phylogenetic analyses of the genus *Gerbillus* did not account for the potential gene tree-species tree incongruence. This incongruence can arise from various factors, either biological (such as incomplete lineage sorting (ILS), introgression, and recombination) or technical

msl	BS-1	BS-2
4	73/70:53/60	–[1]/74:89/76
12	77/81:55/--[2]	–[1]/49:70/82
20	69/76*	76/–[1]*
32	53/50:84/66	--[3]/–[4]:--[3]/--[1]
Reference	44/66:88/59	49/56:96/–[5]

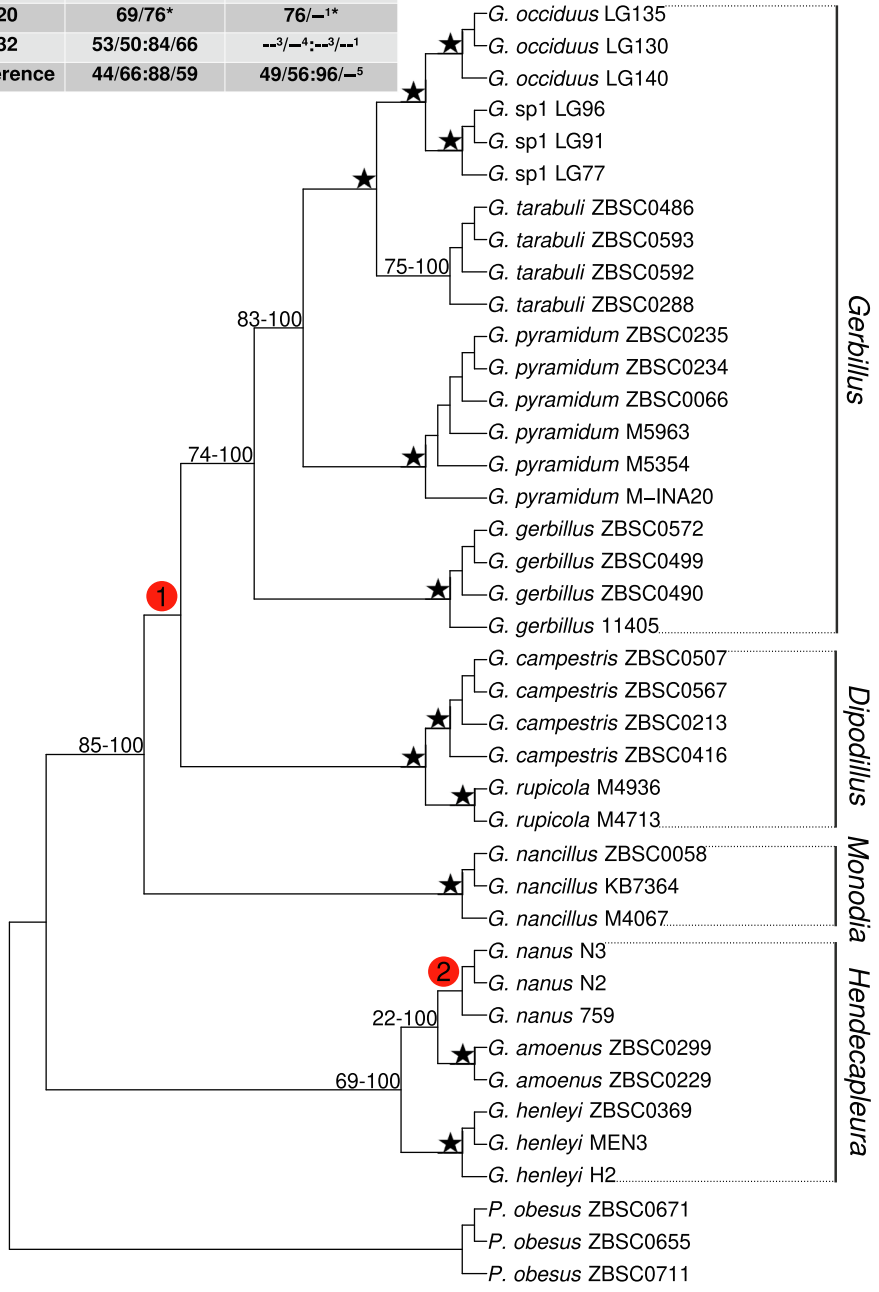

Fig. 2. The dominant topology inferred from SVDquartets analysis of SNP data generated from RAD-seq data for *Gerbillus*. Omitting intraspecific variation, the strongly supported branches in all analyses (96–100%) are marked by stars, whereas the minimal and maximal values of bootstrap are presented for the remaining branches. As in Fig. 1, the two most variable branches are marked ('1' and '2'), and their bootstrap values are given in the table (see Fig. 1 for the detailed description). The lack of bootstrap values in some cases (marked by two dashes) denote the absence of the particular branch, i.e. alternative topology. Each alternative topology is numbered and shown in Fig. D.1. The major clades corresponding to traditionally recognized subgenera are bracketed. The representatives of *Psammomys obesus* were treated as outgroup.

(such as the misidentification of paralogs as orthologs) (reviewed in Bryant and Hahn, 2020). Our analyses revealed that some processes contributing to gene tree discordance operate in the genus *Gerbillus*, but their impact is limited. For instance, we can exclude a substantial amount of ILS or intra-locus recombination because the concatenation analyses yielded a tree consistent with a tree obtained using a coalescence model for the majority of *de novo* generated alignments (Mendes et al., 2019; Mirarab et al., 2016; Warnow, 2015; Figs. 1 and 2). The alternative topologies produced by SVDquartets in a few cases differ from the main topology solely by the placement of two branches, which also exhibited considerable variation in bootstrap support in the concatenation analyses (Figs. 1 and 2). Similarly, the potential usage of paralogs does not pose a problem for the analyses for two reasons. First, considering that species from the genus have large genome sizes (haploid size ≈ 4 Gb according to the c-value database https://www.genomesize.com/ accessed 27.10.2022), which is associated with a large number of potentially paralogous regions, we tested the whole range of clustering thresholds values in ipyrad to find a balance between undersplitting (clustering of non-homologous regions) and oversplitting (splitting of homologous regions into several loci) (McCartney-Melstad et al., 2019). In all cases, the optimal clustering values chosen according to two criteria —the largest mean bootstrap value and the highest number of SNPs—produced concordant results (Figs. 1 and 2). The second reason for the lack of strong effect of potential paralogs assembly on phylogenetic reconstructions is the insensitivity of the reference analyses to changes in the *max_SNPs_locus* parameter. This parameter helps to remove poorly aligned regions with substitution excess, and increasing its value from 0.1 to 0.6 had no effect on the tree topology

Table 4
Summary of quartet sampling scores for the two most variably supported branches in phylogenetic trees obtained from concatenation analyses (see Figs. 1 and 2). All three scores (Quartet Concordance/Quartet Differential/Quartet Informativeness) are shown for the best, in terms of average bootstrap support (avgBS) and number of SNPs (SNP), *de novo* analyses obtained under different values of *min_samples_locus* (msl) and *max_barcode_mismatch* (mbm) parameters (see also table 3). Additionally, quartet scores are provided for reference analyses performed using two values for *max_SNPs_locus*: 0.1 and 0.6.

	Branch '1'		Branch '2'	
msl	mbm = 0	mbm = 1	mbm = 0	mbm = 1
4 (avgBS)	0.42/0.10/0.96	0.14/0.58/0.95	0.67/0.00/0.83	1.00/–/1.00
4 (SNP)	−0.03/0.00/0.90	0.17/0.59/0.89	0.94/0.00/0.99	0.89/0.00/0.99
12 (avgBS)	0.22/0.00/0.90	0.17/0.32/0.90	0.51/0.00/0.81	0.69/0.60/0.85
12 (SNP)	−0.06/0.00/0.95	0.26/0.12/0.90	0.70/0.00/0.94	0.69/0.00/0.90
20 (avgBS, SNP)	0.33/0.04/0.93	0.19/0.34/0.90	1.00/–/0.99	0.70/0.00/0.82
32 (avgBS)	0.10/0.00/0.90	0.40/0.03/0.98	0.14/0.00/0.75	0.39/0.14/0.72
32 (SNP)	0.38/0.00/0.96	0.37/0.14/0.93	0.40/0.06/0.76	0.07/0.00/0.81
Reference (0.1)	0.83/0.00/0.96	0.83/0.20/0.96	0.26/0.00/0.74	0.61/0.14/0.74
Reference (0.6)	0.71/0.53/1.00	0.70/0.08/1.00	0.28/0.00/0.94	0.35/0.00/0.99

Table 5
Results of BFD analyses testing the support of competing species delimitation hypotheses under various prior settings for speciation rate λ and population mutation rate θ. Four species delimitation models were analyzed: (1) full model in which individuals were assigned to species according to the original morphological determination, (2) *Gerbillus occiduus* and putative new species, *G.* sp1, were treated as one species, (3) *G. campestris* and *G. rupicola* were treated as one species, and (4) *G. amoenus* and *G. nanus* were treated as one species. For each hypothesis, the marginal likelihood estimates (MLE), Bayes Factor (BF) and its rank are presented. The model was considered decisively better supported by data when BF > 10.

Model	No. of species	MLE	Rank	BF
Priors: λ = Γ(α = 2.0, β = 200.0), θ = Γ(α = 1.0, β = 2 500.0)				
1	11	−22 244.268	1	–
2	10	−22 372.207	3	255.878
3	10	−22 523.579	4	558.622
4	10	−22 320.613	2	152.690
Priors: λ = Γ(α = 2.0, β = 20.0), θ = Γ(α = 1.0, β = 2 500.0)				
1	11	−22 251.067	1	–
2	10	−22 368.753	3	235.372
3	10	−22 523.062	4	543.990
4	10	−22 319.477	2	136.820
Priors: λ = Γ(α = 2.0, β = 20.0), θ = Γ(α = 1.0, β = 250.0)				
1	11	−21 435.677	1	–
2	10	−21 551.500	2	231.646
3	10	−21 754.365	4	637.376
4	10	−21 570.537	3	269.720
Priors: λ = Γ(α = 2.0, β = 200.0), θ = Γ(α = 1.0, β = 250.0)				
1	11	−21 439.476	1	–
2	10	−21 554.693	2	230.434
3	10	−21 756.949	4	634.946
4	10	−21 574.480	3	270.008

presented here.

The final argument supporting the notion of a limited impact of technical errors and biological processes on phylogenetic reconstruction is the weak effect of changes in the amount of missing data. Generally, the RAD-seq data with large amounts of missing data comprise a mix of rapidly and slowly evolving sites. When removing loci with a high amount of missing data, the mutational spectrum shifts towards slowly evolving sites (Huang and Knowles, 2016). However, this simultaneously reduces the available phylogenetic information (Hovmöller et al., 2013). In *Gerbillus*, we can observe that irrespective of the distribution of the mutational spectrum, the incongruent or reduced phylogenetic signal was consistently limited only to branches '1' and '2' (Fig. 1).

The limited impact of processes causing gene tree-species tree incongruence in *Gerbillus* is responsible for nearly identical results obtained here and in previous studies based exclusively on cyt-b (e.g. Abiadh et al., 2010; Bryja et al., 2022; see also cyt-b tree in Fig. B1) or on cyt-b and one nuclear marker (Ndiaye et al., 2016a). Both concatenation and coalescence analyses consistently recognized four highly to moderately supported clades corresponding to four subgenera (Figs. 1 and 2). This result confirms the placement of *Dipodillus* within *Gerbillus* and the sister position of *Hendecapleura* relative to the other representatives of the genus. However, the relationships between the subgenera remained unresolved mainly due to branch '1' (Figs. 1 and 2), which was also found to be unstable in the Sanger data analyses. For example, previous combined analysis of nuclear and mitochondrial markers by Ndiaye et al. (2016a) suggested the subgenus *Monodia* as sister to *Gerbillus* with poor PP and BS support (Fig. 3 therein), while Bryja et al. (2022) strongly supported the sister relationship between *Dipodillus* and *Gerbillus* based on cyt-b data (Fig. 4 therein), a pattern that also dominated in our analyses.

Branch '1' is a prime example of strong support assigned to conflicting or artifactual relations depending on the method of inference applied. Several factors in our analytical pipeline influenced the support for branch '1'. First, the phylogenetic resolution was affected by the criteria used to determine the best clustering threshold value. McCartney-Melstad et al. (2019) demonstrated that different metrics for determining the best clustering threshold can influence phylogenetic results, recommending the evaluation of multiple criteria to account for possible differences among them. This is evident in our study where the average BS and the number of SNPs yielded different BS values for branch '1' in concatenation analyses: 96% and 44% or 72% and 100%, respectively (Fig. 1). Second, the number of mismatches allowed in the barcode sequence affected the branch support. For example, in one case, the BS for the analysis with no mismatches equaled 96%, compared to 8% when one mismatch was allowed (Fig. 1). The discrepancy arises because demultiplexing raw data with one mismatch generates more reads than demultiplexing with no mismatches permitted, resulting in different numbers of loci carrying a discordant signal for branch '1'. The third factor affecting the support for branch '1' across analyses is the number of missing data. The process of deleting loci might enhance or diminish discordant phylogenetic signal. In concatenation analyses, trees generated based on the matrices with the lowest amount of missing data showed very strong BS for branch '1', while those with an intermediate amount of missing data had as low as 1% of BS (Fig. 1). The fourth factor is the type of phylogenetic analysis used. BI analyses, unlike the ML ones that provided various BS depending on the studied parameters, generated trees with full support for branch '1' (PP = 1.00), regardless of the applied ipyrad parameters. This difference is attributed to the possibility of excessively high PP when using an oversimplified substitution model for phylogenetic inference (Douady et al., 2003; Nylander et al., 2004). Given the size and complexity of concatenated genomic sequences used in this study, this scenario is highly plausible.

The above examples show that BI and PP scores can be misleading in certain situations. The QS method used in this study for concatenated datasets overcomes some of these issues. Despite variation in the QS scores across analyses (Table 4), we can infer the following: (1) branch '1' is poorly supported (QC values never reach 1.00) or even counter-supported in some cases (QC < 0), indicating weak consensus for a dominant relationship (Figs. 1–2), (2) there is moderate to high support for an alternative relationship for branch '1', as shown by QD score values (ranging from 0.00 to 0.59), and (3) the low support for branch '1' is not due to low phylogenetic information (QI score close to 1.00). Interestingly, all generated QD scores supported only one of two possible

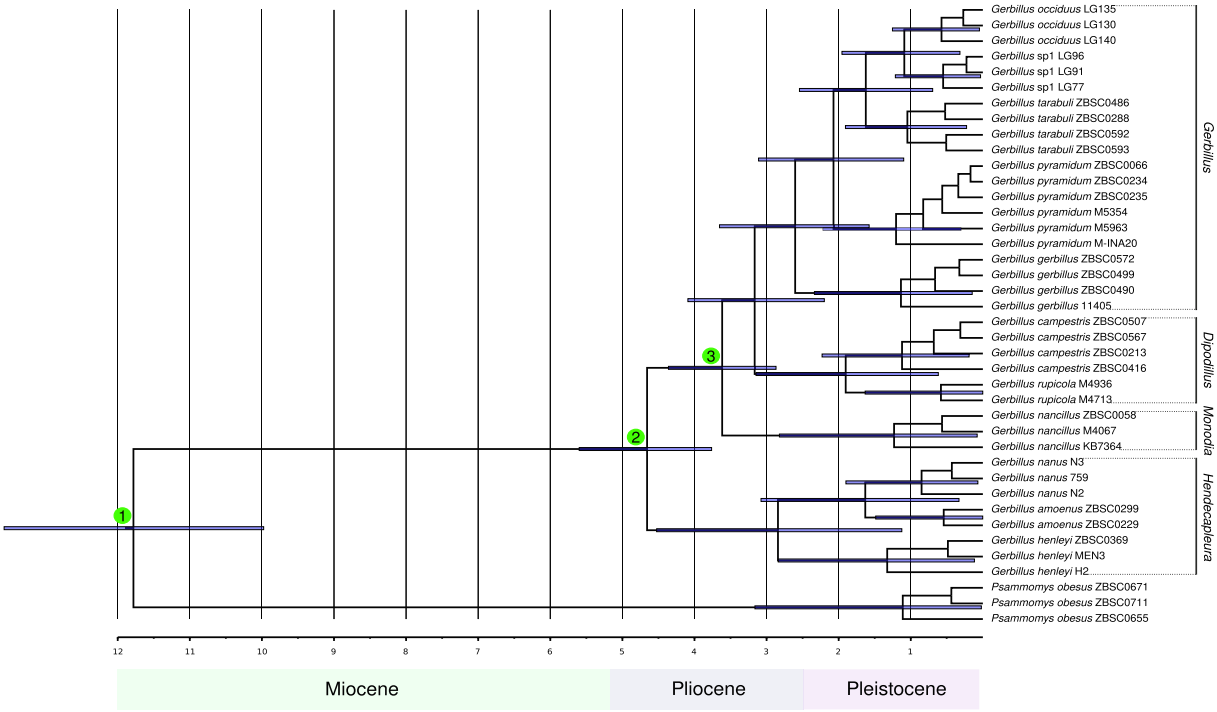

Fig. 3. Maximum clade credibility tree resulting from the BEAST analysis of *Gerbillus* and outgroup (*Psammomys obesus*) under an uncorrelated lognormal molecular clock with the GTR + Γ substitution model. Node ages are represented as mean heights, and node bars indicate the 95% highest posterior density intervals for nodes. Circles with numbers denote secondary calibration points (see Table 2). The major clades corresponding to traditionally recognized subgenera are bracketed. The scale is in million years.

alternative histories for branch '1', in which the subgenus *Dipodillus* is sister to *Monodia,* while *Gerbillus* is sister to *Hendecapleura* (Fig. 1). A similar relationship was found by SVDquartets in one case (see topology 2 in Fig. D1 and Fig. 2), although the clade comprising *Dipodillus* and *Monodia* was sister to the subgenus *Gerbillus*. This incongruent phylogenetic signal displayed by concatenation and coalescence analyses indicates that introgression is a process which likely played a role during the diversification of the main lineages in *Gerbillus*. The distribution of phenotypic and karyological characters in the genus further supports this hypothesis. The subgenera *Hendecapleura* and *Dipodillus* share many chromosomal characters (Aniskin et al., 2006; see also Ndiaye et al., 2016a), while *Gerbillus* and *Hendecapleura* mostly share cranial ones (Lay, 1983; Pavlinov et al., 1990; see also Ndiaye et al., 2016a). Moreover, *G. nancillus,* representing the subgenus *Monodia*, was considered to display intermediate characteristics between species belonging to the subgenera *Gerbillus* and *Dipodillus* (Ellerman, 1941; Ndiaye et al., 2014; Petter, 1968). All these results suggest that an understanding of the diversification of *Gerbillus* lies in an examination of a more complex evolutionary history at a key transition point, which is branch '1', instead of the simple bifurcating model assumed *a priori* by the majority of phylogenetic methods (see, for example, Yu et al., 2014).

4.2. Delimitation analysis corroborates the results of barcoding studies

For two decades, barcoding has been used to evaluate the species diversity in various groups of organisms. In vertebrates, including rodents, the cyt-b gene has become the primary marker used for species identification and delimitation. Although it has successfully unveiled the cryptic diversity in many cases (e.g., Moshtaghi et al., 2016; Rocha et al., 2018), it is not without challenges. A primary concern is the prevalent use of threshold values of genetic divergence to assist species discrimination. Critics point out that no fixed threshold can apply to all studied taxa because of variation in intraspecific divergence, or that discrepancies between genetic divergence and taxonomic assignment can be due to introgression, incomplete lineage sorting or recent speciation (Fregin et al., 2012; Meier et al., 2006; Will and Rubinoff, 2004).

In *Gerbillus*, species boundaries have been primarily determined by comparing genetic distances among cyt-b sequences to the empirical thresholds that Bradley and Baker (2001) provided for different taxonomical levels within Rodentia. For example, comparisons among sister-species showed K2P distances ranging from 2.70% to 19.23% in rodents. Meanwhile, intraspecific comparisons ranged between 0.00% and 6.29% (Bradley and Baker, 2001). These benchmark estimates are useful in *Gerbillus,* where the majority of interspecies distances exceed 10% (e.g., Abiadh et al., 2010; Ndiaye et al., 2016a, 2012). These findings are further corroborated by phylogenetic analyses of cyt-b gene, in which the majority of species identified based on morphology form well differentiated monophyletic groups. However, for certain sister species the given genetic distances did not provide straightforward interpretation. Two species pairs, *G. campestris-G. rupicola* and *G. nanus-G. amoenus,* are particularly noteworthy in this context. *Gerbillus rupicola* was described fairly recently based on a small number of individuals collected in Mali (Granjon et al., 2002). This species is very similar morphologically to *G. campestris,* differing only by molar pattern and karyotype (2n = 52 vs. 2n = 56). Based on phylogenetic analysis of cyt-b and combined analysis of cyt-b and a nuclear marker, *G. rupicola* was found to form a clade with individuals of *G. campestris.* However, its precise position within the clade, sister or nested, remained ambiguous (Ndiaye et al., 2016a; Fig. B1). The results of our phylogenetic analyses clearly placed *G. rupicola* as a sister species to *G. campestris.* The species delimitation analysis provided further confirmation of the specific status of the former (Table 5), despite low mean K2P genetic distance (0.019) estimated for this clade by Ndiaye et al. (2016a). However, there is a limitation of our analysis. The sampling scope is restricted only to *G. campestris* individuals from Morocco and Mauritania, thus omitting a significant portion of the species' distribution (Table 1). Consequently, it remains the possibility that, given a broader sampling, *G. rupicola* might emerge nested within *G. campestris*. Nevertheless, the result of

delimitation analysis indicates that *G. campestris* encompasses either highly divergent populations or morphologically similar species (see also Nicolas et al., 2014). Factors such as chromosome rearrangements and/or changes in chromosome numbers without obvious morphological changes, are likely responsible for recalcitrant relationships in this complex as exemplified by *G. rupicola*.

The species pair *G. nanus-G. amoenus* provides an interesting example of vicariant/allopatric speciation. Initially described from Pakistan, *Gerbillus nanus* was long believed to have a distribution extending from the Balochistan region to the coast of Mauritania by most specialists, whereas *G. amoenus* was thought to be restricted to Libya and Egypt (Ndiaye et al., 2013). Due to their morphological similarity, several hypotheses were proposed to explain this distribution. Cyt-b based analyses supported the hypothesis of a vicariant distribution, with *G. nanus* occurring in Asia and *G. amoenus* in North Africa (Ndiaye et al., 2013; see also Fig. B1). Although both species formed distinct clades, the K2P genetic distance between them was lower than that observed among most species in the genus (0.065). Moreover, this genetic divergence is not accompanied by significant karyotypic changes as observed in *G. campestris-G. rupicola*. However, the delimitation analysis confirmed the results of the cyt-b analysis, strongly supporting the hypothesis that both taxa should be treated as separate species (Table 5).

The last issue that required confirmation through genomic data was the phylogenetic position and status of a taxon considered as a new, hairy-footed gerbil species from the coastal part of the Massa River valley in Western Morocco (*Gerbillus* sp. in Ndiaye et al., 2012). According to cyt-b analysis, this new taxon was placed as a sister to a clade comprising *G. tarabuli* and *G. occiduus,* with a maximal PP (Ndiaye et al., 2012). The latter two sister species which are sympatrically distributed in Morocco, are separated by small genetic distance (K2P = 0.018), but are characterized by a significant number of chromosomal rearrangements (Aniskin et al., 2006). Interestingly, the RAD-seq data provided a different perspective, suggesting that *G. occiduus* is the closest relative of *Gerbillus* sp1 (Figs. 1 and 2). This implies the occurrence of more than one speciation event since the common ancestor of *G. tarabuli* and *G. occiduus,* a finding further supported by the delimitation analysis (Table 5). While there is no karyological data for *Gerbillus* sp1, it is reasonable to speculate that the rapid diversification in this group may have been driven by chromosomal changes. The putative scenario for speciation here may resemble the Kirkpatrick and Barton model (Kirkpatrick and Barton, 2006), where the chromosomal inversions carrying a combination of genes involved in local adaptations are favored by selection if they exhibit low recombination rate. This leads to the fixation of alternative chromosomal rearrangements in different populations, establishing genetic barriers that drive speciation. Moreover, the combination of climatic-driven environmental changes combined with sea level variations in this coastal area of Morocco in the past might have contributed to accelerating the process by enhancing barriers to gene flow (see Ndiaye et al., 2012 for details).

4.3. Genomic data supports the Plio-Pleistocene diversification of *Gerbillus*

The 95% HPD intervals for the calibrated nodes indicated that our estimates were substantially influenced by the priors, making it unclear whether our molecular data set contained significant information about the age of these nodes. One contributing factor might be the distribution of calibration points on the tree. Chazot et al. (2019) observed that using only deep-level fossils resulted in only a slight shift in the posterior distribution compared to the prior. Likely, adding more recent fossils to our analysis would improve the precision of our estimates. Nevertheless, our analysis reaffirmed earlier postulated scenarios regarding the timeline for *Gerbillus* evolution.

Recent reconstructions of climatic history of the Sahara Desert, based on North Atlantic deep-sea sediments, have revealed three main stages of wet-dry oscillations during the last 11 Myr (Crocker et al., 2022). The first stage, spanning from 11.0 to 6.9 Mya, was characterized by the development of arid and dust-producing regions in Africa. It was during this period, in the mid- to late Miocene, that the major lineages of Gerbillinae underwent evolution. For instance, the split between *Sekeetamys* and *Gerbillus* was estimated at 6.48 Mya (95% HPD = 4.56–8.67; Ndiaye et al., 2016a), which is further supported by the fossil record. Several extinct genera, believed to be members of Gerbillinae, have been documented in the late Miocene deposits of Africa, such as *Abudhabia* from Kenya or *Protatera* from Algeria (Winkler et al., 2010). Interestingly, savannahs dominated by C4 grasses expanded extensively during this time, and this expansion has been hypothesized as a major driving force in the evolution of *Gerbillus* (Ndiaye et al., 2016a).

The second period in the evolution of Sahara Desert, which covers the late Miocene to the late Pliocene (5.75–3.5 Mya), was characterized by more humid conditions associated with the development of extensive river systems and lakes, including Lake Chad. During this period, according to our estimates, the subgenus *Hendecapleura* diverged from the rest of the genus (Fig. 3; 95% HPD = 3.76–5.60) and the first fossil of *Gerbillus* was described from Kenya (Winkler et al., 2010). Interestingly, *Hendecapleura* comprises members of the genus adapted to mesic habitats, devoid of typical characteristics of arid-adapted species, such as hairy hind foot soles and large tympanic bullae (Alhajeri and Steppan, 2018). The diversification of *Gerbillus* during this time period supports the idea that the continuous C4-driven expansion of open habitats played a crucial role. The C3 to C4 habitat transition was a gradual long-term trend, which was not directly correlated with climate oscillations in the Plio-Pleistocene period (Crocker et al., 2022).

The third stage of African climate history, spanning from the early Pleistocene to recent times, has been characterized by a shift towards more arid conditions. During the beginning of this period, the split that gave rise to all known lineages within *Gerbillus*, except *Hendecapleura*, occurred (95% HDP = 2.2–4.10; Fig. 3). This crucial split, represented as node '1' in our analyses (Figs. 1 and 2), could have been a pivotal moment in the history of the genus. The decline of ecological and geographical barriers such as river systems in this period that historically kept closely related species apart might have facilitated hybridization and, in consequence, an increase in the likelihood of generating new adaptive genetic variation (Hedrick, 2013). This breakdown of reproductive barriers could resemble processes recently observed in habitats disturbed by human-induced global change (Grabenstein and Taylor, 2018; Vallejo-Marín and Hiscock, 2016).

The emergence of modern species, including *Gerbillus* sp1, was estimated to have occurred 1–2 Mya (Fig. 3), consistent with previous phylogenetic analyses (e.g., Ndiaye et al., 2016a, 2012). This time period, characterized by increasing aridification of the African climate, is hypothesized to have driven extinction and migration in various mammalian taxa, including mega-herbivores such as elephants, rhinos and hippos, as well as hominins (Cerling et al., 2011; Faith et al., 2018). For *Gerbillus*, a taxon with a climatic and ecological niche restricted to arid areas, the Pleistocene represents a period of dynamic evolution, as supported by the extensive fossil record (Tong, 1989) and molecular estimates of the emergence time of modern species (Fig. 3). Diversification processes in *Gerbillus* during this recent period could have been driven by recurrent, rapid and frequent (~6k years) shifts in habitat, from savannah to desert and vice versa, creating a dynamic system where populations were frequently captured in adjacent but isolated habitat patches. Over the last several decades, similar patterns of *Gerbillus* expansion, correlated with aridification and likely with competition release due to the decrease in abundance of other rodents, have been observed. Well-documented examples include the colonization of Burkina Faso and Senegal by *G. henleyi* (Duplantier et al., 1991; Maddalena et al., 1988) and of northern Senegal by *G. tarabuli, G. henleyi* and *G. nigeriae* (Duplantier et al., 1991; Thiam et al., 2008) within the last 30–50 years. The potential evolutionary consequences of these recent changes in distribution and coexistence patterns will undoubtedly be worth observing in the years to come.

5. Conclusions

Despite posing challenges for traditional taxonomists due to the lack of simple morphological markers that can differentiate species, recent phylogenetic studies reveal that modern species of the genus *Gerbillus* are genetically well-differentiated. This likely results from a long evolutionary history dating back to the beginning of the Pliocene, where geographic barriers, climatic and environmental variations as well as fast karyotypic evolution, likely played major roles. The results of this study are in broad agreement with earlier phylogenetic analyses, suggesting that further progress in understanding the evolutionary history of the genus may not necessarily demand more genetic data. Instead, greater emphasis should be placed on including more species that cover the entire distribution range of the genus and obtaining access to well-dated fossil material to better calibrate the *Gerbillus* tree and reduce uncertainty in molecular dating (Parham et al., 2012). We identified only one problematic branch in the resulting phylogenetic tree, leading to diversification of the main lineages in the genus, which appears to be influenced by ancient gene flow among species. The age of this branch coincides with the beginning of aridification in Africa at the Plio-Pleistocene boundary. To corroborate this conclusion, future analyses should include more lineages representing a wider spectra of environmental variation, particularly from the eastern part of Africa (Bryja et al., 2022).

CRediT authorship contribution statement

Marcin Piwczyński: Conceptualization, Formal analysis, Writing – original draft, Writing – review & editing, Data curation, Funding acquisition. **Laurent Granjon:** Resources, Writing – review & editing, Data curation. **Paulina Trzeciak:** Investigation, Writing – review & editing. **José Carlos Brito:** Writing – review & editing. **Madalina Oana Popa:** Investigation. **Mergi Daba Dinka:** Formal analysis. **Nikolas P. Johnston:** Formal analysis, Writing – review & editing. **Zbyszek Boratyński:** Conceptualization, Resources, Writing – original draft, Writing – review & editing, Data curation, Funding acquisition.

Declaration of Competing Interest

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

Acknowledgments

This work was supported by the National Science Center grant no. 2015/18/E/NZ8/00716 to MP and the Portuguese Foundation for Science and Technology grant no. PTDC/BIA-ECO/28158/2017 and National Geographic Society grant no. NGS-53336R-19 projects to ZB. Voucher specimens (and their samples) from Mali, Morocco and Senegal collected by L. Granjon and his colleagues are deposited at the Small mammal Collection of the “Centre de Biologie pour la Gestion des Populations” (https://doi.org/10.15454/WWNUPO).

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ympev.2023.107913.

References

Aberer, A.J., Kobert, K., Stamatakis, A., 2014. Exabayes: Massively parallel bayesian tree inference for the whole-genome era. Mol. Biol. Evol. 31, 2553–2556. https://doi.org/10.1093/molbev/msu236.

Abiadh, A., Chetoui, M., Lamine-Cheniti, T., Capanna, E., Colangelo, P., 2010. Molecular phylogenetics of the genus *Gerbillus* (Rodentia, Gerbillinae): Implications for systematics, taxonomy and chromosomal evolution. Mol. Phylogenet. Evol. 56, 513–518. https://doi.org/10.1016/j.ympev.2010.04.018.

Aghová, T., Kimura, Y., Bryja, J., Dobigny, G., Granjon, L., Kergoat, G.J., 2018. Fossils know it best: Using a new set of fossil calibrations to improve the temporal phylogenetic framework of murid rodents (Rodentia: Muridae). Mol. Phylogenet. Evol. 128, 98–111. https://doi.org/10.1016/j.ympev.2018.07.017.

Alhajeri, B.H., 2021. A geometric morphometric analysis of geographic mandibular variation in the dwarf gerbil *Gerbillus nanus* (Gerbillinae, Rodentia). J. Mamm. Evol. 28, 469–480. https://doi.org/10.1007/s10914-020-09530-9.

Alhajeri, B.H., Steppan, S.J., 2018. A phylogenetic test of adaptation to deserts and aridity in skull and dental morphology across rodents. J. Mammal. 99, 1197–1216. https://doi.org/10.1093/JMAMMAL/GYY099.

Ali, O.A., ORourke, S.M., Amish, S.J., Meek, M.H., Luikart, G., Jeffres, C., Miller, M.R., 2016. RAD capture (Rapture): flexible and efficient sequence-based genotyping. Genetics 202, 389–400. https://doi.org/10.1534/genetics.115.183665.

Aniskin, V.M., Benazzou, T., Biltueva, L., Dobigny, G., Granjon, L., Volobouev, V., 2006. Unusually extensive karyotype reorganization in four congeneric *Gerbillus* species (Muridae: Gerbillinae). Cytogenet. Genome Res. 112, 131–140. https://doi.org/10.1159/000087525.

Barba-Montoya, J., dos Reis, M., Yang, Z., 2017. Comparison of different strategies for using fossil calibrations to generate the time prior in Bayesian molecular clock dating. Mol. Phylogenet. Evol. 114, 386–400. https://doi.org/10.1016/j.ympev.2017.07.005.

Barros, M.I., Brito, J.C., Campos, J.C., Mappes, T., Qninba, A., Sousa, F.V., Boratyński, Z., 2018. The effect of rainfall on population dynamics in Sahara-Sahel rodents. Mammal Res. 63, 485–492. https://doi.org/10.1007/s13364-018-0377-x.

Bensasson, D., Zhang, D.X., Hartl, D.L., Hewitt, G.M., 2001. Mitochondrial pseudogenes: Evolution’s misplaced witnesses. Trends Ecol. Evol. 16, 314–321. https://doi.org/10.1016/S0169-5347(01)02151-6.

Bolger, A.M., Lohse, M., Usadel, B., 2014. Trimmomatic: A flexible trimmer for Illumina sequence data. Bioinformatics 30, 2114–2120. https://doi.org/10.1093/bioinformatics/btu170.

Bonnet, T., Leblois, R., Rousset, F., Crochet, P.A., 2017. A reassessment of explanations for discordant introgressions of mitochondrial and nuclear genomes. Evolution 71, 2140–2158. https://doi.org/10.1111/EVO.13296.

Boratyński, Z., Brito, J.C., Campos, J.C., Cunha, J.L., Granjon, L., Mappes, T., Ndiaye, A., Rzebik-Kowalska, B., Serén, N., 2017. Repeated evolution of camouflage in speciose desert rodents. Sci. Rep. 7, 3522. https://doi.org/10.1038/s41598-017-03444-y.

Bouckaert, R., Vaughan, T.G., Barido-Sottani, J., Duchêne, S., Fourment, M., Gavryushkina, A., Heled, J., Jones, G., Kühnert, D., De Maio, N., Matschiner, M., Mendes, F.K., Müller, N.F., Ogilvie, H.A., Du Plessis, L., Popinga, A., Rambaut, A., Rasmussen, D., Siveroni, I., Suchard, M.A., Wu, C.H., Xie, D., Zhang, C., Stadler, T., Drummond, A.J., 2019. BEAST 2.5: An advanced software platform for Bayesian evolutionary analysis. PLoS Comput. Biol. 15, 1–28. https://doi.org/10.1371/journal.pcbi.1006650.

Bradley, R.D., Baker, R.J., 2001. A test of the genetic species concept: Cytochrome-b sequences and mammals. J. Mammal. 82, 960–973. https://doi.org/10.1644/1545-1542(2001)082<0960:ATOTGS>2.0.CO;2.

Brito, J.C., Godinho, R., Martínez-Freiría, F., Pleguezuelos, J.M., Rebelo, H., Santos, X., Vale, C.G., Velo-Antón, G., Boratyński, Z., Carvalho, S.B., Ferreira, S., Gonçalves, D. V., Silva, T.L., Tarroso, P., Campos, J.C., Leite, J.V., Nogueira, J., Álvares, F., Sillero, N., Sow, A.S., Fahd, S., Crochet, P.-A., Carranza, S., 2014. Unravelling biodiversity, evolution and threats to conservation in the Sahara-Sahel. Biol. Rev. 89, 215–231. https://doi.org/10.1111/brv.12049.

Brito, J.C., Tarroso, P., Vale, C.G., Martínez-Freiría, F., Boratyński, Z., Campos, J.C., Ferreira, S., Godinho, R., Gonçalves, D.V., Leite, J.V., Lima, V.O., Pereira, P., Santos, X., da Silva, M.J.F., Silva, T.L., Velo-Antón, G., Veríssimo, J., Crochet, P.-A., Pleguezuelos, J.M., Carvalho, S.B., 2016. Conservation biogeography of the Sahara-Sahel: additional protected areas are needed to secure unique biodiversity. Divers. Distrib. 22, 371–384. https://doi.org/10.1111/ddi.12416.

Brito, J.C., Del Barrio, G., Stellmes, M., Pleguezuelos, J.M., Saarinen, J., 2021. Drivers of change and conservation needs for vertebrates in drylands: an assessment from global scale to Sahara-Sahel wetlands. Eur. Zool. J. 88, 1103–1129. https://doi.org/10.1080/24750263.2021.1991496.

Bryant, D., Hahn, M.W., 2020. The concatenation question, in: Scornavacca, C., Delsuc, F., Galtier, N. (Eds.), Phylogenetics in the Genomic Era. No commercial publisher, pp. 3.4:1–3.4:23.

Bryja, J., Meheretu, Y., Boratyński, Z., Zeynu, A., Denys, C., Mulualem, G., Welegerima, K., Bryjová, A., Kasso, M., Kostin, D.S., Martynov, A.A., Lavrenchenko, L.A., 2022. Rodents of the Afar Triangle (Ethiopia): geographical isolation causes high level of endemism. Biodivers. Conserv. 31, 629–650. https://doi.org/10.1007/s10531-022-02354-4.

Cerling, T.E., Wynn, J.G., Andanje, S.A., Bird, M.I., Korir, D.K., Levin, N.E., MacE, W., MacHaria, A.N., Quade, J., Remien, C.H., 2011. Woody cover and hominin environments in the past 6 million years. Nature 476, 51–56. https://doi.org/10.1038/nature10306.

Chazot, N., Wahlberg, N., Freitas, A.V.L., Mitter, C., Labandeira, C., Sohn, J.-C., Sahoo, R. K., Seraphim, N., de Jong, R., Heikkilä, M., 2019. Priors and posteriors in Bayesian timing of divergence analyses: The age of butterflies revisited. Syst. Biol. 68, 797–813. https://doi.org/10.1093/sysbio/syz002.

Chifman, J., Kubatko, L., 2014. Quartet inference from SNP data under the coalescent model. Bioinformatics 30, 3317–3324. https://doi.org/10.1093/bioinformatics/btu530.

Crocker, A.J., Naafs, B.D.A., Westerhold, T., James, R.H., Cooper, M.J., Röhl, U., Pancost, R.D., Xuan, C., Osborne, C.P., Beerling, D.J., Wilson, P.A., 2022.

Astronomically controlled aridity in the Sahara since at least 11 million years ago. Nat. Geosci. 15, 671–676. https://doi.org/10.1038/s41561-022-00990-7.
deMenocal, P.B., 1995. Plio-Pleistocene African climate. Science 270, 53–59. https://doi.org/10.1126/science.270.5233.53.
deMenocal, P.B., 2004. African climate change and faunal evolution during the Pliocene-Pleistocene. Earth Planet. Sci. Lett. 220, 3–24. https://doi.org/10.1016/S0012-821X(04)00003-2.
DeSalle, R., Egan, M.G., Siddall, M., 2005. The unholy trinity: Taxonomy, species delimitation and DNA barcoding. Philos. Trans. R. Soc. B Biol. Sci. 360, 1905–1916. https://doi.org/10.1098/rstb.2005.1722.
dos Reis, M., Yang, Z., 2013. The unbearable uncertainty of Bayesian divergence time estimation. J. Syst. Evol. 51, 30–43. https://doi.org/10.1111/j.1759-6831.2012.00236.x.
Douady, C.J., Delsuc, F., Boucher, Y., Doolittle, W.F., Douzery, E.J.P., 2003. Comparison of Bayesian and maximum likelihood bootstrap measures of phylogenetic reliability. Mol. Biol. Evol. 20, 248–254. https://doi.org/10.1093/MOLBEV/MSG042.
Douglas, J., Zhang, R., Bouckaert, R., 2021. Adaptive dating and fast proposals: Revisiting the phylogenetic relaxed clock model. PLoS Comput. Biol. 17, e1008322. https://doi.org/10.1371/journal.pcbi.1008322.
Dubey, S., Michaux, J., Brünner, H., Hutterer, R., Vogel, P., 2009. False phylogenies on wood mice due to cryptic cytochrome-b pseudogene. Mol. Phylogenet. Evol. 50, 633–641. https://doi.org/10.1016/J.YMPEV.2008.12.008.
Duplantier, J.M., Granjon, L., Bâ, K., 1991. Découverte de trois espèces de rongeurs nouvelles pour le Sénégal: un indicateur supplémentaire de la désertification dans le nord du pays. Mammalia 55, 313–315.
Durant, S.M., Wacher, T., Bashir, S., Woodroffe, R., De Ornellas, P., Ransom, C., Newby, J., Abáigar, T., Abdelgadir, M., El Alqamy, H., Baillie, J., Beddiaf, M., Belbachir, F., Belbachir-Bazi, A., Berbash, A.A., Bemadjim, N.E., Beudels-Jamar, R., Boitani, L., Breitenmoser, C., Cano, M., Chardonnet, P., Collen, B., Cornforth, W.A., Cuzin, F., Gerngross, P., Haddane, B., Hadjeloum, M., Jacobson, A., Jebali, A., Lamarque, F., Mallon, D., Minkowski, K., Monfort, S., Ndoassal, B., Niagate, B., Purchase, G., Samaïla, S., Samna, A.K., Sillero-Zubiri, C., Soultan, A.E., Stanley Price, M.R., Pettorelli, N., 2014. Fiddling in biodiversity hotspots while deserts burn? Collapse of the Sahara's megafauna. Divers. Distrib. 20, 114–122. https://doi.org/10.1111/ddi.12157.
Eaton, D.A.R., 2014. PyRAD: Assembly of de novo RADseq loci for phylogenetic analyses. Bioinformatics 30, 1844–1849. https://doi.org/10.1093/bioinformatics/btu121.
Eaton, D.A.R., Ree, R.H., 2013. Inferring phylogeny and introgression using RADseq data: an example from glowering plants (*Pedicularis*: Orobanchaceae). Syst. Biol. 62, 689–706. https://doi.org/10.5061/dryad.bn281.
Eaton, D.A.R., Spriggs, E.L., Park, B., Donoghue, M.J., 2017. Misconceptions on missing data in RAD-seq phylogenetics with a deep-scale example from flowering plants. Syst. Biol. 66, 399–412. https://doi.org/10.1093/sysbio/syw092.
Edgar, R.C., 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. 32, 1792–1797. https://doi.org/10.1093/nar/gkh340.
Ellerman, J., 1941. The families and genera of living rodents. Hayman & Holt, London.
Faith, J.T., Rowan, J., Du, A., Koch, P.L., 2018. Plio-Pleistocene decline of African megaherbivores: No evidence for ancient hominin impacts. Science 362, 938–941. https://doi.org/10.1126/science.aau2728.
Fernández-Mazuecos, M., Mellers, G., Vigalondo, B., Sáez, L., Vargas, P., Glover, B.J., 2018. Resolving recent plant radiations: Power and robustness of genotyping-by-sequencing. Syst. Biol. 67, 250–268. https://doi.org/10.1093/sysbio/syx062.
Fregin, S., Haase, M., Olsson, U., Alström, P., 2012. Pitfalls in comparisons of genetic distances: A case study of the avian family Acrocephalidae. Mol. Phylogenet. Evol. 62, 319–328. https://doi.org/10.1016/J.YMPEV.2011.10.003.
Gouy, M., Guindon, S., Gascuel, O., 2010. SeaView version 4: A multiplatform graphical user interface for sequence alignment and phylogenetic tree building. Mol. Biol. Evol. 27, 221–224. https://doi.org/10.1093/molbev/msp259.
Grabenstein, K.C., Taylor, S.A., 2018. Breaking barriers: Causes, consequences, and experimental utility of human-mediated hybridization. Trends Ecol. Evol. 33, 198–212. https://doi.org/10.1016/J.TREE.2017.12.008.
Granjon, L., Aniskin, V.M., Volobouev, V., Sicard, B., 2002. Sand-dwellers in rocky habitats: a new species of *Gerbillus* (Mammalia: Rodentia) from Mali. J. Zool. 256, 181–190. https://doi.org/10.1017/S0952836902000213.
Grummer, J.A., Bryson, R.W., Reeder, T.W., 2014. Species delimitation using bayes factors: Simulations and application to the sceloporus scalaris species group (Squamata: Phrynosomatidae). Syst. Biol. 63, 119–133. https://doi.org/10.1093/sysbio/syt069.
Grzywacz, A., Trzeciak, P., Wiegmann, B.M., Cassel, B.K., Pape, T., Walczak, K., Bystrowski, C., Nelson, L., Piwczyński, M., 2021. Towards a new classification of Muscidae (Diptera): a comparison of hypotheses based on multiple molecular phylogenetic approaches. Syst. Entomol. 46, 508–525. https://doi.org/10.1111/syen.12473.
Hedrick, P.W., 2013. Adaptive introgression in animals: examples and comparison to new mutation and standing variation as sources of adaptive variation. Mol. Ecol. 22, 4606–4618. https://doi.org/10.1111/MEC.12415.
Hipp, A.L., Eaton, D.A.R., Cavender-Bares, J., Fitzek, E., Nipper, R., Manos, P.S., 2014. A framework phylogeny of the American oak clade based on sequenced RAD data. PLoS One 9, e93975. https://doi.org/10.1371/journal.pone.0093975.
Hovmöller, R., Lacey Knowles, L., Kubatko, L.S., 2013. Effects of missing data on species tree estimation under the coalescent. Mol. Phylogenet. Evol. 69, 1057–1062. https://doi.org/10.1016/J.YMPEV.2013.06.004.
Huang, H., Knowles, L.L., 2016. Unforeseen consequences of excluding missing data from next-generation sequences: Simulation study of RAD sequences. Syst. Biol. 65, 357–365. https://doi.org/10.1093/sysbio/syu046.
Inoue, J., Donoghue, P.C.J., Yang, Z., 2010. The impact of the representation of fossil calibrations on Bayesian estimation of species divergence times. Syst. Biol. 59, 74–89. https://doi.org/10.1093/sysbio/syp078.
Kirkpatrick, M., Barton, N., 2006. Chromosome inversions, local adaptation and speciation. Genetics 173, 419–434. https://doi.org/10.1534/genetics.105.047985.
Lataste, F., 1881. Diagnoses de mammifères nouveaux d'Algerie. 3. *Gerbillus simoni*, n. sp. Le Nat. 63, 497–500.
Lataste, F., 1882. Mammifères nouveaux d'Algérie. Le Nat. 16, 126–127.
Lay, D., 1983. Taxonomy of the genus *Gerbillus* (Rodentia: Gerbillinae) with comments on the applications of generic and subgeneric names and an annotated list of species. Zeitschrift für Säugetierkd. im Auftrage der Dtsch. Gesellschaft für Säugetierkd. e.V 48, 329–354.
Leaché, A.D., Fujita, M.K., Minin, V.N., Bouckaert, R.R., 2014. Species delimitation using genome-wide SNP Data. Syst. Biol. 63, 534–542. https://doi.org/10.1093/sysbio/syu018.
Lu, J.Y., Shao, W., Chang, L., Yin, Y., Li, T., Zhang, H., Hong, Y., Percharde, M., Guo, L., Wu, Z., Liu, L., Liu, W., Yan, P., Ramalho-Santos, M., Sun, Y., Shen, X., 2020. Genomic repeats categorize genes with distinct functions for orchestrated regulation. Cell Rep. 30, 3296–3311.e5. https://doi.org/10.1016/j.celrep.2020.02.048.
Maddalena, T., Sicard, B., Tranier, M., Gautun, J.-C., 1988. Note sur la présence de *Gerbillus henleyi* (de Winton, 1903) au Burkina Faso. Mammalia 52, 282–285.
McCartney-Melstad, E., Gidiş, M., Shaffer, H.B., 2019. An empirical pipeline for choosing the optimal clustering threshold in RADseq studies. Mol. Ecol. Resour. 19, 1195–1204. https://doi.org/10.1111/1755-0998.13029.
Meier, R., Shiyang, K., Vaidya, G., Ng, P.K.L., 2006. DNA barcoding and taxonomy in Diptera: a tale of high intraspecific variability and low identification success. Syst. Biol. 55, 715–728. https://doi.org/10.1080/10635150600969864.
Mendes, F.K., Hahn, M.W., 2018. Why concatenation fails near the anomaly zone. Syst. Biol. 67, 158–169. https://doi.org/10.1093/sysbio/syx063.
Mendes, F.K., Livera, A.P., Hahn, M.W., 2019. The perils of intralocus recombination for inferences of molecular convergence. Philos. Trans. R. Soc. B Biol. Sci. 374, 20180244. https://doi.org/10.1098/rstb.2018.0244.
Miller, M.A., Pfeiffer, W., Schwartz, T., 2010. Creating the CIPRES Science Gateway for inference of large phylogenetic trees., in: Proceedings of the Gateway Computing Environments Workshop (GCE). IEEE, New Orleans, LA, pp. 1–8. https://doi.org/10.1109/GCE.2010.5676129.
Mirarab, S., Bayzid, M.S., Warnow, T., 2016. Evaluating summary methods for multilocus species tree estimation in the presence of incomplete lineage sorting. Syst. Biol. 65, 366–380. https://doi.org/10.1093/sysbio/syu063.
Moritz, C., Cicero, C., 2004. DNA barcoding: Promise and pitfalls. PLoS Biol. 2, e354. https://doi.org/10.1371/journal.pbio.0020354.
Moshtaghi, S., Darvish, J., Mirshamsi, O., Mahmoudi, A., 2016. Cryptic species diversity in the genus *Allactaga* (Rodentia: Dipodidae) at the edge of its distribution range. Folia Zool. 65, 142–147. https://doi.org/10.25225/fozo.v65.i2.a9.2016.
Musser, G., Carleton, M., 2005. Superfamily Muroidea. In: Wilson, D., Reeder, D. (Eds.), Mammal Species of the World: A Taxonomic and Geographic Reference. John Hopkins University Press, Baltimore, pp. 894–1531.
Ndiaye, A., Bâ, K., Aniskin, V., Benazzou, T., Chevret, P., Konečný, A., Sembène, M., Tatard, C., Kergoat, G.J., Granjon, L., 2012. Evolutionary systematics and biogeography of endemic gerbils (Rodentia, Muridae) from Morocco: an integrative approach. Zool. Scr. 41, 11–28. https://doi.org/10.1111/j.1463-6409.2011.00501.x.
Ndiaye, A., Shanas, U., Chevret, P., Granjon, L., 2013. Molecular variation and chromosomal stability within *Gerbillus nanus* (Rodentia, Gerbillinae): Taxonomic and biogeographic implications. Mammalia 77, 105–111. https://doi.org/10.1515/mammalia-2012-0039.
Ndiaye, A., Hima, K., Dobigny, G., Sow, A., Dalecky, A., Bâ, K., Thiam, M., Granjon, L., 2014. Integrative taxonomy of a poorly known Sahelian rodent, *Gerbillus nancillus* (Muridae, Gerbillinae). Zool. Anzeiger - A J. Comp. Zool. 253, 430–439. https://doi.org/10.1016/j.jcz.2014.03.004.
Ndiaye, A., Chevret, P., Dobigny, G., Granjon, L., 2016a. Evolutionary systematics and biogeography of the arid habitat-adapted rodent genus *Gerbillus* (Rodentia, Muridae): a mostly Plio-Pleistocene African history. J. Zool. Syst. Evol. Res. 54, 299–317. https://doi.org/10.1111/jzs.12143.
Ndiaye, A., Tatard, C., Stanley, W., Granjon, L., 2016b. Taxonomic hypotheses regarding the genus *Gerbillus* (Rodentia, Muridae, Gerbillinae) based on molecular analyses of museum specimens. Zookeys 566, 145–155. https://doi.org/10.3897/zookeys.566.7317.
Near, T.J., Sanderson, M.J., 2004. Assessing the quality of molecular divergence time estimates by fossil calibrations and fossil–based model selection. Philos. Trans. R. Soc. London. Ser. B Biol. Sci. 359, 1477–1483. https://doi.org/10.1098/rstb.2004.1523.
Nicolas, V., Ndiaye, A., Benazzou, T., Souttou, K., Delapre, A., Denys, C., 2014. Phylogeography of the North African dipodil (Rodentia: Muridae) based on cytochrome-b sequences. J. Mammal. 95, 241–253. https://doi.org/10.1644/13-MAMM-A-241/2/JMAMMAL-95-2-241.
Nokelainen, O., Brito, J.C., Scott-Samuel, N.E., Valkonen, J.K., Boratyński, Z., 2020. Camouflage accuracy in Sahara-Sahel desert rodents. J. Anim. Ecol. 89, 1658–1669. https://doi.org/10.1111/1365-2656.13225.
Nylander, J.A.A., Ronquist, F., Huelsenbeck, J.P., Nieves-Aldrey, J.L., 2004. Bayesian phylogenetic analysis of combined data. Syst. Biol. 53, 47–67. https://doi.org/10.1080/10635150490264699.
Parham, J.F., Donoghue, P.C.J., Bell, C.J., Calway, T.D., Head, J.J., Holroyd, P.A., Inoue, J.G., Irmis, R.B., Joyce, W.G., Ksepka, D.T., Patané, J.S.L., 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., Müller, J., Smith, K.T., Theodor, J.M.,

Warnock, R.C.M., Benton, M.J., 2012. Best practices for justifying fossil calibrations. Syst. Biol. 61, 346–359. https://doi.org/10.1093/sysbio/syr107.
Pausata, F.S.R., Gaetani, M., Messori, G., Berg, A., Maia de Souza, D., Sage, R.F., DeMenocal, P.B., 2020. The greening of the Sahara: Past changes and future implications. One Earth 2, 235–250. https://doi.org/10.1016/j.oneear.2020.03.002.
Pavlinov, I., 2008. A review of phylogeny and classification of Gerbillinae (Mammalia: Rodentia). Moscow University Publishing, Moscow.
Pavlinov, I.J., Dubrovsky, A.Y., Potabova, E.G., Rossolimo, O.L., 1990. Gerbillids of the wild word. Nauka Publication, Moscow.
Pease, J.B., Brown, J.W., Walker, J.F., Hinchliff, C.E., Smith, S.A., 2018. Quartet Sampling distinguishes lack of support from conflicting support in the green plant tree of life. Am. J. Bot. 105, 385–403. https://doi.org/10.1002/AJB2.1016.
Petter, F., 1968. Rodentia: Gerbillinae (excluding the genera *Tatera* and *Gerbillurus*). In: Meester, J. (Ed.), Preliminary Indentification Manual for African Mammals. Smithsonian Institution, Washington.
Petter, F., 1975. Subfamily Gerbillinae. Part 6. 3. In: Meester, J., Setzer, H. (Eds.), The Mammals of Africa: An Identification MAnual. Smithsonian Institution, Washington, pp. 7–12.
Piwczyński, M., Trzeciak, P., Popa, M.O., Pabijan, M., Corral, J.M., Spalik, K., Grzywacz, A., 2021. Using RAD seq for reconstructing phylogenies of highly diverged taxa: A test using the tribe Scandiceae (Apiaceae). J. Syst. Evol. 59, 58–72. https://doi.org/10.1111/jse.12580.
Rambaut, A., Drummond, A.J., Xie, D., Baele, G., Suchard, M.A., 2018. Posterior summarization in Bayesian phylogenetics using Tracer 1.7. Syst. Biol. 67, 901–904. https://doi.org/10.1093/sysbio/syy032.
Rannala, B., Yang, Z., 2020. Species delimitation., in: Scornavacca, C., Delsuc, F., Galtier, N. (Eds.), Phylogenetics in the genomic era. https: //hal.inria.fr/PGE/hal-02536468.
Reaz, R., Bayzid, M.S., Rahman, M.S., 2014. Accurate phylogenetic tree reconstruction from quartets: A heuristic approach. PLoS One 9, e104008. https://doi.org/10.1371/journal.pone.0104008.
Richly, E., Leister, D., 2004. NUMTs in sequenced eukaryotic genomes. Mol. Biol. Evol. 21, 1081–1084. https://doi.org/10.1093/MOLBEV/MSH110.
Rocha, R.G., Duda, R., Flores, T., Rossi, R., Sampaio, I., Mendes-Oliveira, A.C., Leite, Y.L. R., Costa, L.P., 2018. Cryptic diversity in the *Oecomys roberti* complex: revalidation of *Oecomys tapajinus* (Rodentia, Cricetidae). J. Mammal. 99, 174–186. https://doi.org/10.1093/JMAMMAL/GYX149.
Schuster, M., Duringer, P., Ghienne, J.-F., Vignaud, P., Mackaye, H.T., Likius, A., Brunet, M., 2006. The age of the Sahara desert. Science 311, 821–821. https://doi.org/10.1126/science.1120161.
Shafer, A.B.A., Peart, C.R., Tusso, S., Maayan, I., Brelsford, A., Wheat, C.W., Wolf, J.B.W., 2017. Bioinformatic processing of RAD-seq data dramatically impacts downstream population genetic inference. Methods Ecol. Evol. 8, 907–917. https://doi.org/10.1111/2041-210X.12700.
Stamatakis, A., 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics 30, 1312–1313. https://doi.org/10.1093/bioinformatics/btu033.
Steppan, S.J., Schenk, J.J., 2017. Muroid rodent phylogenetics: 900-species tree reveals increasing diversification rates. PLoS One 12, e0183070. https://doi.org/10.1371/journal.pone.0183070.
Swezey, C.S., 2009. Cenozoic stratigraphy of the Sahara, Northern Africa. J. African. Earth Sci. 53, 89–121. https://doi.org/10.1016/j.jafrearsci.2008.08.001.
Swofford, D., 2003. PAUP*. Phylogenetic Analysis Using Parsimony (*and Other Methods). Version 4.
Takahashi, T., Nagata, N., Sota, T., 2014. Application of RAD-based phylogenetics to complex relationships among variously related taxa in a species flock. Mol. Phylogenet. Evol. 80, 137–144. https://doi.org/10.1016/j.ympev.2014.07.016.
Thiam, M., Bâ, K., Duplantier, J., 2008. Impacts of climatic changes on small mammal communities in the Sahel (West Africa) as evidenced by owl pellet analysis. African Zool. 43, 135–143. https://doi.org/10.1080/15627020.2008.11657230.
Tong, H., 1989. Origine et évolution des Gerbillidae (Mammalia, Rodentia) en Afrique du Nord. Mémoires la Société géologique Fr. 155, 1–120.
Tranier, M., Julien-Laferriere, D., 1990. A propos de petites gerbilles du Niger et du Tchad (Rongeurs, Gerbillidae, *Gerbillus*). Mammalia 54, 451–456. https://doi.org/10.1515/mamm.1990.54.3.451.
Tripp, E.A., Tsai, Y.-H.-E., Zhuang, Y., Dexter, K.G., 2017. RADseq dataset with 90% missing data fully resolves recent radiation of *Petalidium* (Acanthaceae) in the ultra-arid deserts of Namibia. Ecol. Evol. 7, 7920–7936. https://doi.org/10.1002/ece3.3274.
Vallejo-Marín, M., Hiscock, S.J., 2016. Hybridization and hybrid speciation under global change. New Phytol. 211, 1170–1187. https://doi.org/10.1111/NPH.14004.
Wagner, C.E., Keller, I., Wittwer, S., Selz, O.M., Mwaiko, S., Greuter, L., Sivasundar, A., Seehausen, O., 2013. Genome-wide RAD sequence data provide unprecedented resolution of species boundaries and relationships in the Lake Victoria cichlid adaptive radiation. Mol. Ecol. 22, 787–798. https://doi.org/10.1111/mec.12023.
Ward, D., 2009. The biology of deserts, 1st ed. Oxford University Press, Oxford.
Warnock, R.C.M., Parham, J.F., Joyce, W.G., Lyson, T.R., Donoghue, P.C.J., 2015. Calibration uncertainty in molecular dating analyses: There is no substitute for the prior evaluation of time priors. Proc. R. Soc. B Biol. Sci. 282, 20141013. https://doi.org/10.1098/rspb.2014.1013.
Warnow, T., 2015. Concatenation analyses in the presence of incomplete lineage sorting. PLoS Curr. 1–10 https://doi.org/10.1371/currents.tol.8d41ac0f13d1abedf4c4a59f5d17b1f7.
Will, K.W., Rubinoff, D., 2004. Myth of the molecule: DNA barcodes for species cannot replace morphology for identification and classification. Cladistics 20, 47–55. https://doi.org/10.1111/J.1096-0031.2003.00008.X.
Winkler, A., Denys, C., Avery, D., 2010. Rodentia, in: Werdelin, L., Sanders, W. (Eds.), Cenozoic Mammals of Africa. University of California Press, Berkeley, pp. 263–305.
Yu, Y., Dong, J., Liu, K.J., Nakhleh, L., 2014. Maximum likelihood inference of reticulate evolutionary histories. PNAS 111, 16448–16453. https://doi.org/10.1073/PNAS.1407950111/SUPPL_FILE/PNAS.1407950111.SAPP.PDF.
Zomer, R.J., Xu, J., Trabucco, A., 2022. Version 3 of the global aridity index and potential evapotranspiration database. Sci. Data 9, 409. https://doi.org/10.1038/s41597-022-01493-1.