THERYA, 2025, Vol. 16(1):29-39 DOI:10.12933/therya-25-6138 ISSN 2007-3364

Are the Heteromyidae paraphyletic? Molecular phylogenetics of extant geomyoid rodents

Enrique P. Lessa*, and Andrés Parada

Departamento de Ecología y Evolución, Facultad de Ciencias, Universidad de la República. Iguá 4225, Montevideo 11400, Uruguay. E-mail: enrique.lessa@gmail.com (EPL); andrespara@gmail.com (AP).
*Corresponding author: https://orcid.org/0000-0001-9750-4093.

The superfamily Geomyoidea includes the Geomyidae (pocket gophers: 7 genera, 42 species) and Heteromyidae (pocket mice, kangaroo mice, and kangaroo rats: 5 genera, 69 species). Analyses of both morphological and molecular data have confirmed the monophyly of the superfamily relative to other rodents but have cast doubts on the reciprocal monophyly of Geomyidae and Heteromyidae. The latter are recovered as paraphyletic in some, but not all molecular phylogenies, with low to moderate support for critical nodes. To test alternative hypotheses of geomyoid phylogenetic relationships, we searched NCBI databases and assembled four datasets: a) the 13 protein-coding mitochondrial DNA genes (7 genera, 13 species); b) the mitochondrial 12s, 16s, and COX1 loci (7 genera, 43 species); and two datasets of ultraconserved elements (UCEs; 6 genera, 9 species): c) one with 3,991 loci, allowing for up to 2 unrepresented taxa per locus, and d) another one reduced to 1,750 UCEs (with no missing data). In all cases, beavers were included as outgroups. Maximum likelihood analysis of both mitochondrial datasets were equivocal regarding heteromyid paraphyly, as support for the critical nodes was very low. In contrast, the trees obtained from UCE loci with both Maximum Likelihood and a multi-species coalescent method (wASTRAL) indicated that pocket gophers were sister to pocket mice (100 % bootstrap, local Posterior Probability of 1), to the exclusion of kangaroo rats, which formed a second, strongly supported clade (100 % bootstrap, local Posterior Probability of 1; Figure 3). These preliminary findings are consistent with the hypothesis that Heteromyidae are paraphyletic relative to Geomyidae. The inclusion of additional taxa (*e. g.*, kangaroo mice) in the analyses is required to confirm these results.

La superfamilia Geomyoidea incluye Geomyidae (tuzas: 7 géneros, 42 especies) y Heteromyidae (ratones de abazones, ratones y ratas canguro: 5 géneros, 69 especies). Los análisis de datos tanto morfológicos como moleculares han confirmado la monofilia de la superfamilia con respecto a otros roedores, pero arrojaron dudas sobre la monofilia recíproca de Geomyidae y Heteromyidae. Estos últimos se recuperan como parafiléticos en algunas, pero no en todas, las filogenias moleculares, con un apoyo bajo a moderado para los nodos críticos. Para poner a prueba hipótesis alternativas sobre las relaciones filogenéticas de los geomioideos, buscamos en las bases de datos del NCBI y reunimos cuatro dos conjuntos de datos: a) los 13 genes del ADN mitocondrial que codifican proteínas (7 géneros, 13 especies); b) los loci mitocondriales 12s, 16s y COX1 (7 géneros, 43 especies); y dos conjuntos de datos de elementos ultraconservados (UCEs; 6 géneros, 9 especies): c) uno con 3,991 loci, permitiendo hasta 2 taxa no representados por locus; y d) otro con 1,750 UCEs (sin datos faltantes). En todos los casos, los castores fueron incluidos como grupo externo. El análisis de Máxima Verosimilitud de loci mitocondriales concatenados recuperó a Geomyidae y Heteromyidae como taxones hermanos, pero el apoyo de "bootstrap" para la monofilia de este último fue sólo del 55 %. Por el contrario, los árboles obtenidos de loci UCE, tanto con Máxima Verosimilitud como con un método coalescente para múltiples especies (wASTRAL) indicaron que las tuzas eran hermanas de los ratones de abazones (valor de "bootstrap" de 100%, Probabilidad Posterior local de 1), con exclusión de las ratas canguro, que formaban un segundo clado fuertemente apoyado (valor de "bootstrap" de 100%, Probabilidad Posterior local de 1). Estos hallazgos preliminares son consistentes con la hipótesis de que la familia Heteromyidae es parafilética con respecto a Geomyidae. Se requiere la inclusión de taxones adicionales (*e. g.*, ratones canguro) en los análisis para confirmar estos resultados.

Keywords: Geomyidae; Geomyoidea; mitogenomes; systematics; ultraconserved elements.

Introduction

Our understanding of the diversity and phylogenetic relationships of rodents has been greatly impacted by DNA sequencing data. Molecular phylogenetic studies have helped support alternative hypotheses based on morphological and paleontological data, while also suggesting new phylogenetic relationships and uncovering hidden diversity among morphologically cryptic species. The integration of molecular and morphological data has resulted in the recognition of three major clades of extant rodents, namely the squirrel-related clade (suborder Sciuromorpha), the guinea pig-related clade (suborder Hystricomorpha = Ctenohystrica), and the mouse-related clade (reviewed by Fabre *et al.* 2015; D'Elía *et al.* 2019; see also Swanson *et al.* 2019; Bangs and Steppan 2022).

The mouse-related clade includes three major groups, typically recognized as suborders: 1) the families Anomaluridae, Pedetidae, and Zenkerellidae (Suborder Anomaluromorpha); 2) a diverse assemblage of nine families, grouped into the superfamilies Muroidea and Dipodoidea (Suborder Myomorpha; and 3) a clade including the families Castoridae, Heteromyidae, and Geomyidae (Suborder Castorimorpha). These three suborders may be treated as infraorders if the mouse-related clade is recognized as the Suborder Supramyomorpha (D'Elía *et al.* 2019; Flynn *et al.* 2019).

Here, we focus on the phylogenetic relationships of extant Castorimorpha (or Castorimorphi), an assemblage of some 113 species comprising the Castoridae (2 species of beavers), the Geomyidae (42 species of pocket gophers), and the Heteromyidae, subdivided into the subfamilies Heteromyinae (17 species of spiny pocket mice), Perognathinae (30 species of pocket mice), and Dipodomyinae (2 species of kangaroo mice and 20 species of kangaroo rat, (Table 1; Mammal Diversity Database 2023). A. E. Wood (1955) first proposed the suborder Castorimorpha to include the Castoridae and the fossil family Eutypomidae. An unanticipated result of molecular phylogenetic studies was the sister taxon relationship between Castoridae and the superfamily Geomyoidea Weber 1904 (*cf*. Wahlert 1985), that includes the families Heteromyidae and Geomyidae (Huchon *et al.* 1999; Murphy *et al.* 2001; Adkins *et al.* 2003; DeBry 2003; but see DeBry and Sagel 2001). Carleton and Musser (2005) tentatively accepted the inclusion of Geomyidae and Heteromyidae within Castorimorpha, an arrangement that was subsequently supported by additional molecular phylogenetic studies (*e. g.* Montgelard *et al.* 2008; Blanga-Kanfi *et al.* 2009; Fabre *et al.* 2012; Fabre *et al.* 2015; Swanson *et al.* 2019; but see Veniaminova *et al.* 2007).

Recognition of the close relationships of Geomyidae and Heteromyidae (superfamily Geomyoidea) predates their inclusion in the Castorimorpha (*e. g.* Wood 1931) and has received strong, confirmatory support from molecular phylogenetics (reviewed by Fabre *et al.* 2015). Both families share the following characteristics: "shortened incisive foramina (less than half diastema length); posterior maxillary foramen closed; cranially, the premaxillae extend further posteriorly than the nasal bones; there is a parapterygoid fossa and a bulla at the back of foramen ovale; the cheek teeth tend to be high crowned and lophodont, with stylar cusps well developed" (Flynn *et al.* 2008). The reciprocal monophyly of Geomyidae and Heteromyidae, however, has been questioned on both morphological and molecular grounds. In particular, paleontological studies relying on similarity in dental morphology have suggested that both extant pocket gophers and heteromyids might have been derived from ancestors similar to pocket mice that may be placed within the Heteromyidae (*e. g.*, Lindsay 1972; Korth (1994).

However, Wahlert (1985: pp. 14 to 15) examined the cranial morphology of both extant and fossil geomyoids and identified 18 synapomorphies in support of the monophyly of pocket gophers, as well as seven synapomorphies in support of the monophyly of extant heteromyids. Analyses of myological characters also identified several synapomorphies in support of the reciprocal monophyly of Geomyidae and Heteromyidae (Ryan 1989). In contrast, Brylski (1990) examined the development and adult anatomy of carotid circulation of geomyoids and found that both pocket gophers and *Heteromys* lost the stapedial arteries in adults, in contrast with other heteromyids. If the absence of stapedial arteries found in pocket gophers and *Heteromys* is a derived condition, it suggests either heteromyid paraphyly or parallel evolution of carotid circulation in these taxa (Brylski 1990). In general, morphological analyses of relationships among extant geomyoids have favored reciprocal monophyly of at least crown Heteromyidae and Geomyidae (see discussion in Wahlert 1985; Korth (1994; Hafner *et al.* 2007).

Early molecular phylogenetic analyses of higher-level relationships among rodents (*e. g.* (Huchon *et al.* 1999; Adkins *et al.* 2001; Adkins *et al.* 2003) typically included few castorimorphs. DeBry (2003) included three geomyoids in his analysis of rodent phylogenetics and obtained moderate support for a *Thomomys* + *Chaetodipus* clade to the exclusion of *Dipodomys*. In contrast, Montgelard *et al.* (2008) recovered Geomyidae (represented by three genera) and Heteromyidae (represented by *Dipodomys* and *Heteromys*) as reciprocally monophyletic.

Using sequence data from three mitochondrial loci (12s ribosomal RNA, 16s ribosomal RNA, and COX1), Hafner *et al.* (2007) examined phylogenetic relationships within the Heteromyidae, with multiple species per genus and four genera of Geomyini as outgroup taxa. This study offered moderate to strong support for each of the three heteromyid subfamilies (Figure 1). Hafner *et al.* (2007) also reviewed the relevant literature and argued in favor of the reciprocal monophyly of Heteromyidae and Geomyidae, although no outgroup to the Geomyoidea was included in their study.

The analysis of Fabre *et al.* (2012) included a total of 87 species of castorimorphs (31 geomyids, 54 heteromyids, and the 2 extant beaver species), representing a major advance in taxonomic representation in a multilocus analysis of higher-level relationships among rodents. Fabre *et al.* (2012) analyzed DNA sequence data from six mitochondrial and five nuclear loci, recovering both the castorimorph clade and the sister-taxon relationships of Castoridae and Geomyoidea. However, Heteromyidae were recovered as paraphyletic relative to Geomyidae and, within heteromyids, the Perognathinae were not recovered as monophyletic (Figure 1).

Although Fabre *et al.* (2012) used a total of 11 loci, these were unevenly represented across taxa. The number of mitochondrial loci represented in at least one species per genus ranged from 1 to 4, whereas the corresponding range for nuclear loci was 0 to 5. Furthermore, the nodes that are key to the paraphyly of heteromyids relative to geomyids were not well supported (Figure 1). In their review of rodent phylogenetics, Fabre *et al.* (2015) combined the loci used by Montgelard *et al.* (2008) and Blanga-Kanfi et al. (2009); in the resulting phylogeny (their Figure 2.1), *Dipodomys* was recovered as sister to three geomyid genera, forming a clade that excluded *Heteromys*, but this relationship was weakly supported.

Within the Geomyidae, molecular phylogenetic analyses strongly support the classical view that the tribes Thomomyini (genus *Thomomys*) and Geomyini (six genera, see Table 1) are sister taxa (Spradling *et al.* 2004; Fabre *et al.* 2012).

Recently, datasets with numerous nuclear loci have enabled some authors to revisit these phylogenetic hypotheses. A study based on 2,213 ultraconserved elements (UCEs, Swanson *et al.* 2019) found high support for both castorimorphs and geomyoids, but the relationships within the latter (represented by one pocket gopher, one kangaroo rat, and one pocket mouse species) were unresolved. Upham *et al.* (2019) used a "backbone and patch" approach to examine available DNA sequence data to build a phylogeny of all extant mammalian species. Their study included sequence data for 109 castorimorphs, represented by up to 24 nuclear and three mitochondrial loci. Their phylogeny (*e. g.* Upham *et al.* 2019: Supplementary Material 3 Data) also shows Heteromyidae as paraphyletic, but pocket gophers appear as sister to pocket mice, a result different from that of Fabre *et al.* (2012). Again, support for heteromyid paraphyly is moderate (91 % bootstrap support; Figure 1). Table 2 provides a summary of the taxonomic and genetic coverage of selected molecular phylogenetic studies.

In summary: 1) the suborder Castorimorpha (or infraorder Castorimorphi) has been strongly supported by molecular data and includes the family Castoridae and the Geomyoidea as a strongly supported superfamily; 2) among the latter, the Geomyidae have consistently been recovered as a monophyletic group; 3) some, but not all morphological and, especially, several molecular phylogenetic analyses suggest that Heteromyidae are paraphyletic relative to Geomyidae; however, 4) trees that suggest paraphyly are not consistent in the identity of the closest relatives of Geomyidae among Heteromyidae, and support for the key nodes is low to moderate in these phylogenies.

Recent genomic projects, including Zoonomia (https://zoonomiaproject.org/) and the California Conservation Genomics Project (https://www.ccgproject.org/) have made available high quality genomes of several castorimorphs. Additionally, the annotation of vertebrate mitogenomes from raw sequences has been greatly simplified by new software (*e. g.* Meng *et al.* 2019), making it possible to improve the representation for mitochondrial sequences. In this study, we take advantage of full genomes and other genomic datasets to revisit the issue of phylogenetic relationships of families and subfamilies within the Geomyoidea, with the goal of reassessing the hypothesis of heteromyid paraphyly. To this end, we use two main types of genomic data: a mitogenomic dataset that includes complete codons of the 13 protein-coding genes encoded in the mitochondrial genome; and a dataset of UCEs that combines the loci sequenced by Swanson *et al.* (2019) with their orthologs obtained by mining five additional available castorimorph genomes. We also reanalyzed the data from Hafner *et al.* (2007) after adding sequences of the two extant beaver species (*Castor canadensis* and *C. fiber*) as outgroups.

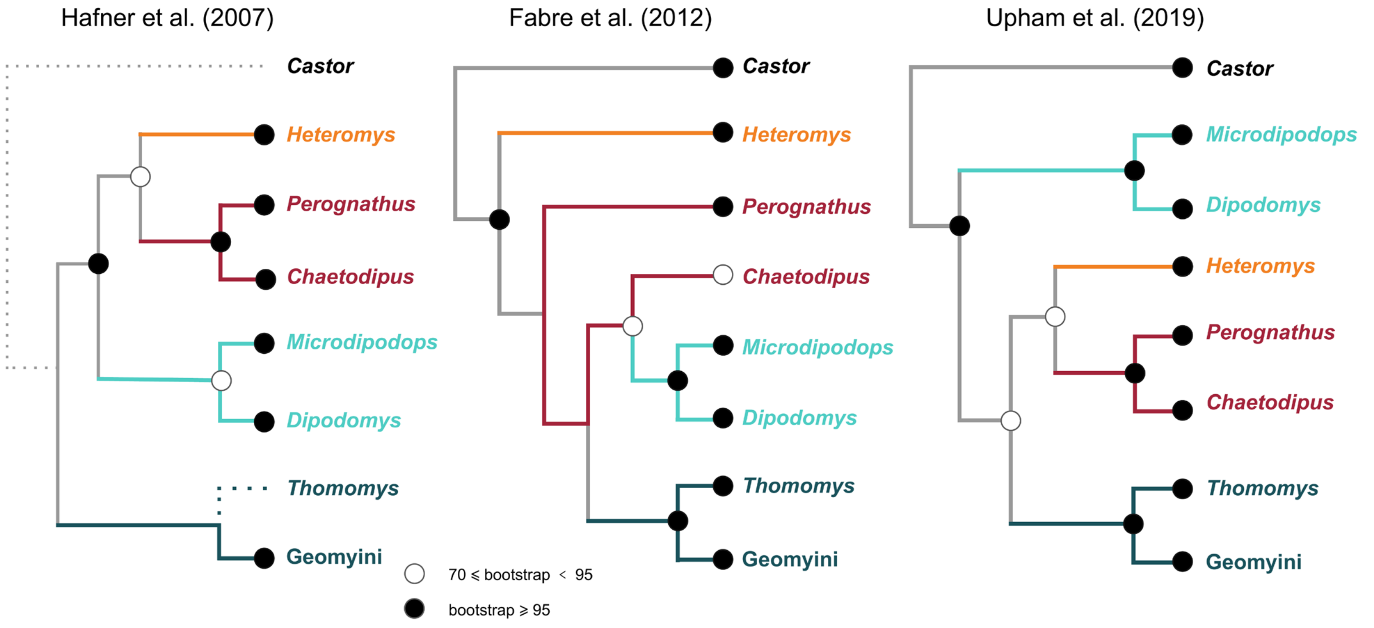

Figure 1. Three hypotheses of phylogenetic relationships among castorimorphs. The currently recognized subfamilies of Heteromyidae are the Heteromyinae (orange), Perognathinae (maroon) and Dipodomyinae (turquoise). Circles denote support for the corresponding nodes in the trees. Hafner *et al.* (2007) argue for reciprocal monophyly of Geomyidae and Heteromyidae; this study did not include *Castor* or *Thomomys*, but their positions in the tree (dotted lines) are implied in the author's discussion. Fabre *et al.* (2012) recovered Heteromyidae as paraphyletic relative to Geomyidae, with the latter being sister to an assemblage of Dipodomyinae + *Chaetodipus*. Upham *et al.* (2019) recovered Heteromyid paraphyly, with Geomyidae sister to pocket mice (Heteromyinae + Perognathinae).

Materials and methods

Mitogenomes. The protein-coding sequences of annotated mitogenomes were downloaded from NCBI for the following species: *Castor canadensis* (NC_033912.1; Lok *et al*. 2017); *Castor fiber birulai* (OQ731787.1); *Chaetodipus penicillatus* (NC_068810.1); *Dipodomys spectabilis* (NC_068808.1); *Dipodomys merriami* (NC_068807.1); *Geomys pinetis* (NC_069016.1); and *Thomomys bottae* (CM063115.1). Additional sets of protein-coding loci encoded by the mitochondrial genome were assembled de novo from RNAseq or genomic reads available in the Sequence Read Archive (SRA) of NCBI: *Perognathus longimembris* (SRR1846776; Kozak *et al.* 2024); *Chaetodipus baileyi* (SRR1633402); *Chaetodipus intermedius* (SRR12430174); *Heteromys desmarestianus* (SRR12430176); *Dipodomys ordii* (SRR1646418); and *Dipodomys stephensi* (SRR14572526). In these cases, mitogenomes were obtained with the software MitoZ 3.6 (Meng *et al.* 2019). The final dataset included the 13 protein-coding loci for 13 taxa, with the following missing data: MT-ND1 (Mitochondrially Encoded NADH Dehydrogenase 1 was not recovered for the two species of beavers and *Chaetodipus intermedius*, and MT-ND2 was not recovered for the two species of beavers, *C. intermedius*, and *Perognathus longimembris*.

Not all mitogenomic data are linked to museum vouchers. As a check of the species identity of the sequences included in our study, we used BLAST (Camacho *et al*. 2009) on the NCBI site to assess the similarity of the CYTB and COX1 DNA sequences of each of our taxa against the NCBI Nucleotide database. In all cases, the highest similarity scores were to sequences of the same named species reported by a peer-reviewed publication.

The strength of the mitogenomic data set described above is the inclusion of all protein-coding mitochondrial loci at the expense of taxonomic representation. To evaluate the impact of taxonomic sampling in this dataset, we assembled a second mitochondrial DNA dataset by adding the two species of beavers to the taxon-rich matrix reported by Hafner *et al.* (2007). The sequences of the 12s and 16s ribosomal loci and of COX1 for beavers were obtained from the aforementioned mitogenomes. This expanded dataset included representatives from all genera and 43 species, with COX1 missing for *Cratogeomys estor* and 12s missing for *Heterogeomys dariensis*.

Ultraconserved elements (UCEs). We used the UCEs reported by Swanson *et al.* (2019) for *Castor canadensis*, *Heteromys oasicus*, *Dipodomys ordii*, and *Cratogeomys planiceps* (BioSample/genome accessions at NCBI: SAMN10715119, SAMN10715105, GCA_000151885.2_Dord_2.0, and SAMN10715121, respectively). In addition, we used the software phyluce v1.7.3 (Faircloth 2016) to mine the reference genomes of *Thomomys bottae* (Museum of Vertebrate Zoology: voucher MVZ:Mamm:240275; reference genome: mThoBot1.0.p; Genbank accession: GCA_024803745.1; Voss *et al.* 2024), *Dipodomys ordii* (Dord_2.0, GCF_000151885.1); *Dipodomys merriami* (MVZ:Mamm:240054, mDipMer1.0.p, GCA_024711535.1), *Dipodomys spectabilis* (Auburn University:Male_0828, ASM1905484v1, GCA_019054845.1); Harder *et al.* 2022), *Dipodomys stephensi* (Broad Institute:BS19, DipSte_v1_BIUU, GCA_004024685.1), and *Perognathus longimembris* (San Diego Zoo Wildlife Alliance:PPM17, ASM2315922v1, GCA_023159225.1; Kozak *et al.* 2024).

The number of UCE loci recovered per taxon were as follows: *Castor canadensis* (3,700), *Dipodomys merriami* (3,643), *D. ordii* (3,947), *D. spectabilis* (3,953), *D. stephensi* (3,966), *Heteromys oasicus* (2,359), *Perognathus longimembris* (3,856), *Cratogeomys planiceps* (2,852), and *Thomomys bottae* (3,837). The final, full UCE dataset, allowing for up to two missing taxa per locus, included 3,991 loci.

Analyses were also carried out separately on a reduced UCE dataset of the 1,750 loci that were available for all taxa.

Alignments and phylogenetic analyses. Alignments of each of the 4 datasets were done using MAFFT v7.520 (Katoh and Standley 2013).

Phylogenetic analysis for concatenated data was carried out in IQ-TREE 2.2.5 (Minh *et al.* 2020), separately for mitogenomes, the three mitochondrial loci used by Hafner *et al.* (2007), and the full (3,991 loci) or reduced (1,750 loci) UCE data. For each data set, each locus was treated as a separate partition, for which model selection was carried out independently with ModelFinder within IQ-TREE (Kalyaanamoorthy *et al.* 2017). ModelFinder uses a greedy strategy that first considers a full partition model and sequentially merges partitions, so long as there is no significant loss in model fit.

To circumvent high consumption of RAM memory, node support with the UCEs dataset was assessed considering a single partition (with TVM+F+I+R5, a transversion model with unequal base frequencies estimated empirically, a fraction of invariant sites, and five categories of rates to allow for rate heterogeneity among sites– selected as the best model with ModelFinder, using the Bayesian Information Criterion).

Node support for each of the four datasets was carried out with 1,000 pseudo-replicas with an ultrafast bootstrap approximation with Maximum Likelihood (ML) in IQ-TREE; the best ML tree was annotated with percent bootstrap support for each node.

Independently for each UCE locus, model selection was carried out with ModelFinder as described above, and a Maximum Likelihood (ML) gene tree was recovered. Using those trees as input, the species tree was obtained using a multispecies coalescent-based method implemented in wASTRAL-unweighted v1.16.3.4 (Zhang *et al.* 2018; Zhang and Mirarab 2022). Local posterior probability (local PP) which is the probability that a branch is true given the set of gene trees (Sayyari and Mirarab 2016) was used as an estimate of support for each node.

Results

Mitochondrial phylogenies. The dataset for the 13 protein-coding loci for 13 taxa consisted of 11,491 sites, of which 5,061 were parsimony-informative and 1,009 were singletons. The ML phylogeny recovered from this dataset is presented in Figure 2. Within the superfamily Geomyoidea, the bootstrap consensus tree shows Heteromyidae and Geomyidae as sister taxa, *i. e.*, the two families are recovered as reciprocally monophyletic. However, support for reciprocal monophyly is very low (55 % bootstrap value). This phylogeny shows support for monophyly of the Geomyidae, as expected from previous phylogenetic efforts. As for internal relationships within Heteromyidae, the monophyly of *Dipodomys* and of the Perognathinae (*Perognathus* + *Chaetodipus*) are well supported.

Our second mitochondrial dataset (that adds the 12s, 16s, and COX1 loci of beavers to the geomyoid matrix produced by Hafner *et al.* 2007, for a total of 43 taxa) included 4,251 sites, of which 1,166 were parsimony-informative and 448 were singletons. The Maximum Likelihood phylogeny based on this data set provides a complementary perspective on the mitochondrial phylogeny of castorimorphs, with strong taxon representation at the expense of additional loci. The resulting tree (Figure 2) shows strong support (100 % bootstrap) for Geomyidae, and the subfamily Heteromyinae. Support for the Dipodomyinae and Perognathinae is moderate (89 % and 93 % bootstrap support, respectively). The Heteromyidae are recovered as paraphyletic relative to Geomyidae. Specifically, the tree includes a clade formed by Perognathinae, Dipodomyinae and Geomyidae to the exclusion of Heteromyinae, but with low bootstrap support (81 %). The sister taxon relationship between Dipodomyinae and Geomyinae is even less supported (bootstrap <50 %).

Additional information for each locus, the partition scheme, and selected substitution models, as well as details of the results, are provided in the Supplementary materials.

Phylogeny based on Ultraconserved elements (UCEs). The 3,991 UCE loci had an average length of 725 sites (range: 368 to 2,290), for a total of 2,895,348 sites, of which 139,516 were parsimony-informative and 350,453 were singletons (additional information of each locus is provided in Supplementary Table 1). The phylogeny based on the full UCE data using a multi-species coalescent method (as implemented in wASTRAL) is presented in Figure 3. Kangaroo rats (four species of *Dipodomys*), pocket mice (*Heteromys oasicus* + *Perognathus longimembris*), and pocket gophers (*Cratogeomys planiceps* + *Thomomys bottae*) are strongly supported as monophyletic units (local Posterior Probability of 1). In addition, pocket gophers are recovered as sister to pocket mice (local PP of 1) to the exclusion of kangaroo rats. A phylogeny based on a concatenation of the UCE loci analyzed by Maximum Likelihood (results not shown) recovers all these clades, including the sister taxon relationship between pocket gophers and pocket mice (to the exclusion of kangaroo rats) with 100 % bootstrap support.

Both the multi-species coalescent (wASTRAL) and concatenated (Maximum Likelihood) analyses recover the pair formed by *D. ordii* and *D. stephensi* with local PP of 1. However, wASTRAL places *D. merriami* as sister to this pair, but support for this relationship is low (local PP = 0.67);

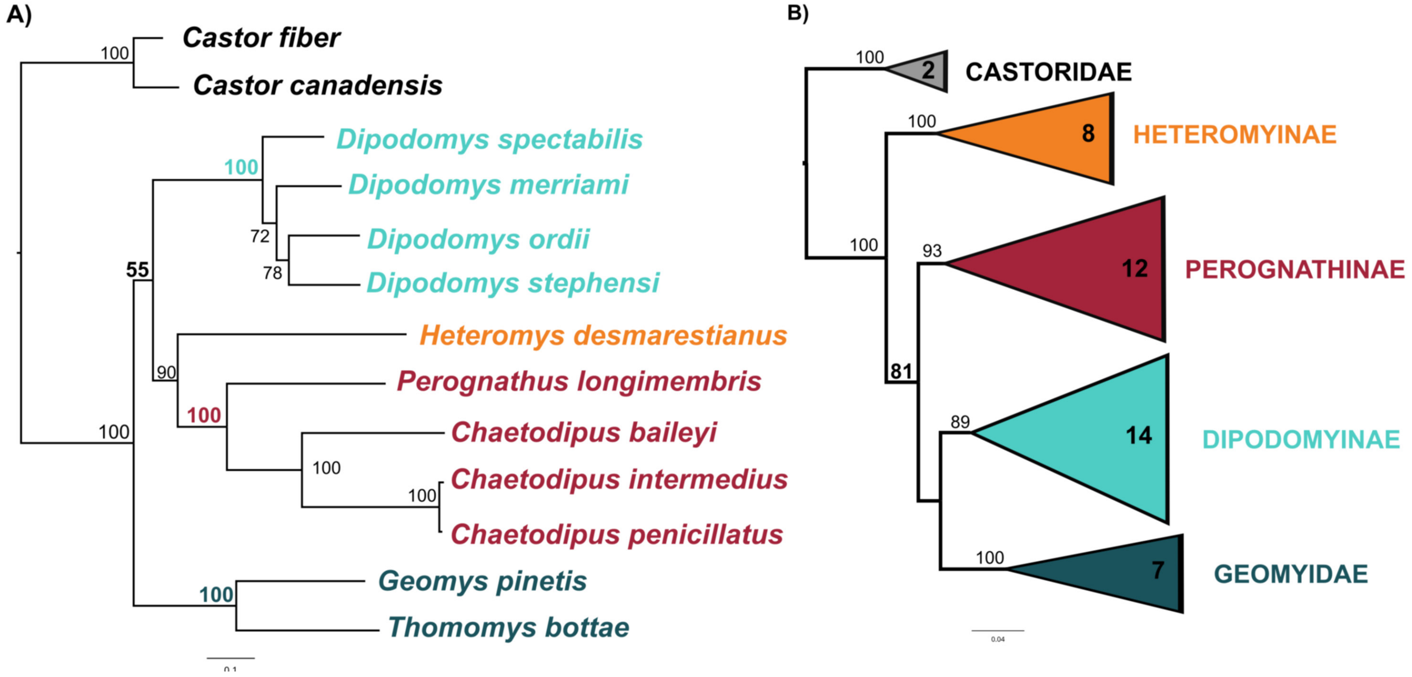

Figure 2. Phylogenies of castorimorph rodents based on mitochondrial DNA sequences; bootstrap values for nodes shown if > 50 %. A) Maximum Likelihood phylogeny based on the 13 protein-coding loci encoded by the mitochondrial genome. B) Reanalysis of Hafner *et al.* (2007) data for 12s, 16s, and COX1, after including the 2 beaver species as outgroups to the Geomyoidea. The number of species included for each family or subfamily is shown within the corresponding triangles.

Table 1. Classification and diversity of extant castorimorphs (suborder Castorimorpha, in Wilson and Reeder (2005); Fabre *et al.* (2015); infraorder Castorimorphi in D'Elía *et al.* (2019). The number of species per genus was obtained from the Mammal Diversity Database (2023), v1.12.1, released 30 Jan 2024.

Superfamily	Family	Subfamily	Tribe	Genus	Number of species	Common name
	Castoridae			*Castor*	2	beavers
Geomyoidea	Heteromyidae	Heteromyinae		*Heteromys*	17	spiny pocket mice
		Perognathinae		*Perognathus*	10	pocket mice
				Chaetodipus	20	pocket mice
		Dipodomyinae		*Microdipodops*	2	kangaroo mice
				Dipodomys	20	kangaroo rats
Geomyoidea	Geomyidae		Geomyini	*Cratogeomys*	7	pocket gophers
				Geomys	13	"
				Heterogeomys	7	"
				Orthogeomys	1	"
				Pappogeomys	1	"
				Zygogeomys	1	"
			Thomomyini	*Thomomys*	12	"

in contrast, Maximum Likelihood recovers *D. spectabilis* as sister to the *D. ordii* + *D. stephensi* pair with strong (local PP = 0.98) support.

The reduced UCE dataset included 1,750 loci with an average length of 729 sites, for a total of 1,276,436 sites, of which 68,826 were parsimony-informative and 143,039 were singletons. Using both wASTRAL and Maximum Likelihood analyses, this dataset produced topologies that were identical and with local PP of 1 and 100 bootstrap support, respectively, for the monophyly of kangaroo rats, pocket mice, and pocket gophers; the sister-taxon relationships of pocket mice and pocket gophers to the exclusion of kangaroo rats was also well supported in both cases (local PP = 1; 100 % bootstrap).

As for the relationships among kangaroo rats obtained from the reduced UCE dataset, each of the methods recovered the same topology obtained with the full dataset. However, whereas support for the topology obtained with wASTRAL increased to a local PP of 0.76, support for the Maximum Likelihood phylogeny dropped from 98 % to 64 %.

Additional information of each locus, the partition scheme, and selected substitution models, as well as details of the results, are provided in the Supplementary materials.

Discussion

Phylogenetics of Geomyoidea: where do we stand? The monophyly of the family Geomyidae is strongly supported by morphological studies (*e. g.*, Wahlert 1985) and, unsurprisingly, by virtually all molecular phylogenies (see Introduction; Figure 1), including our reanalysis of the data from Hafner *et al.* (2007), the tree obtained from mitogenomic data (Figure 2), as well as our phylogenies based on UCEs (Figure 3), with the caveat that the latter have only one Geomyini and one Thomomyini. The reciprocal monophyly of these two tribes is also fully supported by molecular phylogenies (*e. g.*, Spradling *et al.* 2004; Upham *et al.* 2019).

Previous molecular phylogenetic studies have shown increasing support for the subfamilies of Heteromyidae (Figure 1). The Dipodomyinae (*Dipodomys* + *Microdipodops*) are well supported in the studies of Fabre *et al.* (2012) and Upham *et al.* (2019), and the Heteromyinae (*Heteromys*, including *Liomys* as a junior synonym) are well supported in Hafner *et al.* (2007) and subsequent studies. Although Fabre *et al.* (2012) did not recover a monophyletic Perognathinae (*Perognathus* + *Chaetodipus*), support for the monophyly of this subfamily is strong in both the mitochondrial study of Hafner *et al.* (2007) and the analysis that combines nuclear and mitochondrial loci of Upham *et al.* (2019). Our mitochondrial genomic data (Figure 2) provide strong support for the monophyly of the Perognathinae, with the caveat that taxonomic representation is limited.

In sum, there is good support for the monophyly of each of the four major groups of geomyoids, namely the family Geomyidae and the subfamilies Heteromyinae, Perognathinae, and Dipodomyinae. The relationships among them, however, remain problematic. As discussed above, reciprocal monophyly of Geomyidae and Heteromyidae is recovered in some analyses, including our mitogenomic tree (Figure 2); however, support for that topology is very low (55 % bootstrap). Collapsing the key node results in a trichotomy, with branches leading to Dipodomyinae, Perognathinae + Heteromyinae, and Geomyidae. As such, neither our reanalysis of the data from Hafner *et al.* (2007) with beavers as an outgroup nor our mitogenomic analyses support the reciprocal monophyly of the two families.

As discussed in the Introduction, the alternative hypothesis (paraphyly of Heteromyidae relative to Geomyidae) has been recovered in some, but not all molecular phylogenies, generally with low support (see Figure 1). Our UCE data provide strong support for the sister taxon relationship between pocket mice (Heteromyinae + Perognathinae) and pocket gophers (Geomyidae), to the exclusion of Dipodomyinae (Figure 3), with the caveat that taxonomic sampling is limited to eight geomyoids and one beaver as an outgroup. In particular, the genera *Chaetodipus* and *Microdipodops* are not represented in the data. Also missing are several genera of Geomyini, although there is little question about the monophyly of Geomyidae. Interestingly, a topology suggesting that pocket mice (Heteromyinae + Perognathinae) are sister to Geomyidae is recovered in both maximum likelihood analyses of concatenated UCE data and in the coalescent-based phylogenies based on UCE loci. Furthermore, the topology is similar to the one obtained, with less support, by Upham *et al.* (2019), a study that is much stronger in taxonomic representation (see Table 2). It should be noted, however, that, although Upham *et al.* (2019) used four mitochondrial and 24 nuclear loci of Castorimorpha, the representation of most of the nuclear loci is limited to a few species.

Taken collectively, our and earlier studies seem to indicate that mitochondrial sequences cannot resolve relationships among the subfamilies of Heteromyidae and the family Geomyidae. Nuclear data, alone or combined with mitochondrial data, have much greater promise, as shown by Upham *et al.* (2019) and by our analysis of UCE data, both of which lean in support of heteromyid paraphyly, with pocket mice sister to pocket gophers. These results are provisional and will have to be re-examined with a good combination of taxonomic representation and large numbers of loci. We discuss some options to achieve that goal in the next section.

Molecular phylogenetics of Geomyoidea in the genomic era. In principle, PCR amplification and sequencing of the loci used by Upham *et al.* (2019) can provide a more thorough representation of genes across taxa, as required to resolve higher-level relationships within the Geomyoidea. Targeted enrichment protocols, combined with next generation sequencing, provide several alternatives to locus-by-locus sequencing. For example, Bangs and Steppan (2022) developed a set of probes to obtain > 400 loci specifically selected for rodent phylogenetics; importantly, these loci have been shown to be useful at different levels of divergence within rodents. Radseq or ddRADseq is yet another effective method for sampling a few thousand genes (Peterson *et al.* 2012), but the approach appears to work best to examine variation within and between closely related species (*e. g.*, Tomasco *et al.* 2024). Exome capture can potentially provide sequences of the coding regions of thousands of loci, and the approach is adaptable to the needs of each project, especially at the population level (Bi *et al.* 2013). For higher-level relationships, UCE tetrapod probes can effectively capture up to 5,000 loci (McCormack *et al.* 2012; Esselstyn *et al.* 2017; McLean *et al.* 2019; Swanson *et al.* 2019).

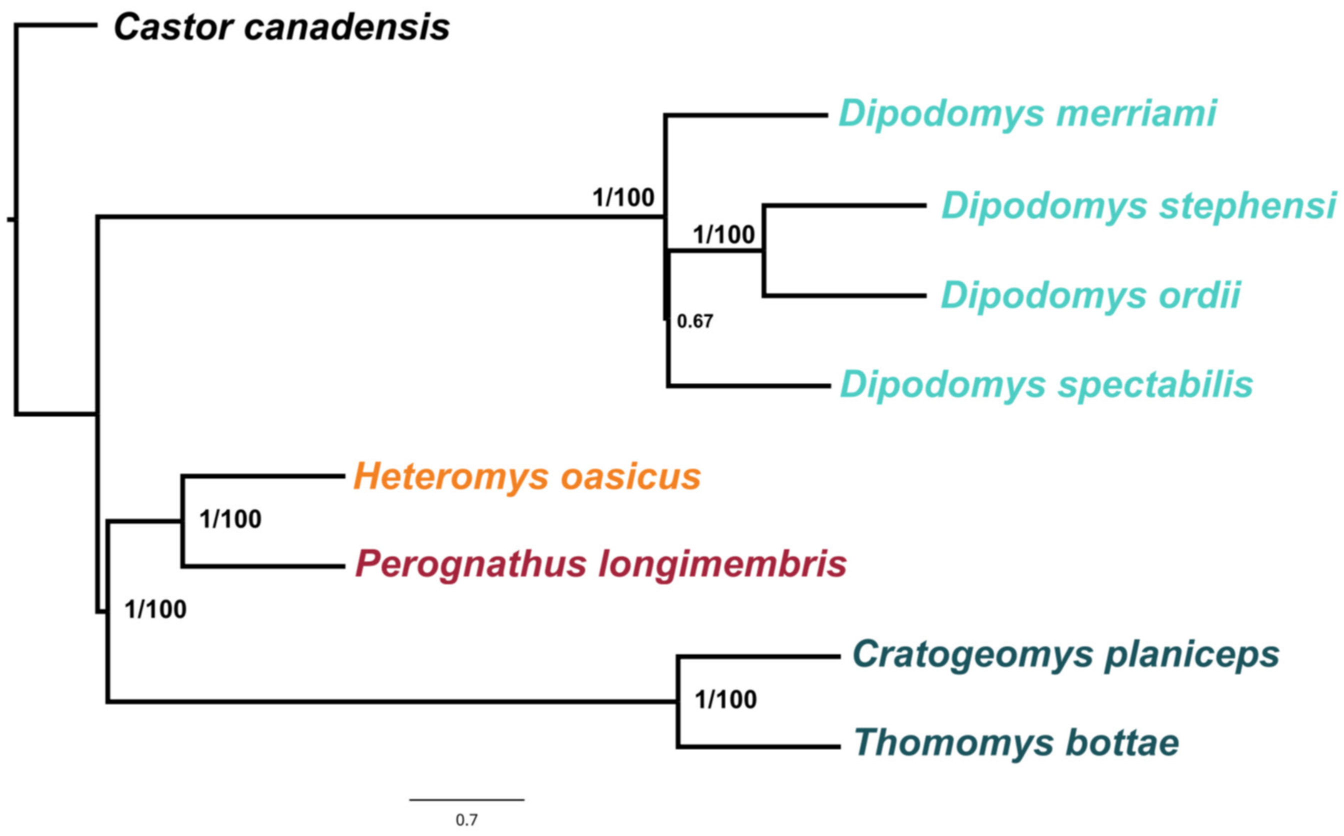

Figure 3. Phylogeny of castorimorph rodents using ultraconserved elements, obtained with wASTRAL and the 3,991 loci dataset. Support values are shown for the local posterior probability (local PP) from wASTRAL and analysis of 1,000 ultrafast bootstrap Maximum Likelihood pseudo-replicates (%). Only local PP support is shown for a node within *Dipodomys*, as this relationship is not recovered with Maximum Likelihood (see Results).

Table 2. Taxonomic and genetic coverage of selected studies of phylogenetic relationships of castorimorphs based on DNA sequences. The number of species per genus (in parenthesis) is taken from the Mammal Diversity Database (2023) in the first column, and from each of the studies in the remaining columns.

Family and genus	Spradling *et al.* 2004	Hafner *et al.* 2007	Montgelard *et al.* 2008	Blanga-Kanfi *et al.* 2009	Fabre et al. 2012	Fabre *et al.* 2015	Upham *et al.* 2019
Castoridae							
Castor (2)			1	1	2	1	2
Geomyidae							
Cratogeomys (7)	5	1	1		8	1	7
Geomys (12)	2		1		8	1	10
Heterogeomys (17)	3				4		
Orthogeomys (1)	1	1			1		11
Pappogeomys (1)	1	1			2		2
Zygogeomys (1)	1	1			1		1
Thomomys (12)	6		1	1	7	1	12
Heteromyidae							
Dipodomys (20)		12	1	1	19	1	21
Microdipodops (2)		2			2		2
Heteromys[a] (17)		8	1		9	1	15
Chaetodipus (20)	1	7			15		17
Perognathus (10)		4			9		9
Total (112)	**20**	**37**	**6**	**3**	**87**	**6**	**109**
Number of loci							
mitochondrial	2	3	2		6		4
nuclear	2		6	6	5	10	24
Total	**4**	**3**	**8**	**6**	**11**	**10**	**28**

a. Includes *Liomys* as a junior synonym.

Any of the loci recovered by these methods can be mined from the growing number of available genomes to supplement sequencing efforts. In the case of UCEs, software is available to extract them from genomes (Faircloth 2016), and the resulting data can be combined with those from UCE studies, as illustrated by our own UCE dataset (see also Swanson *et al.* 2019; Parada *et al.* 2021). The number of available genomes will continue to increase and taxonomic representation will improve along the way. Even full genomes, however, cannot resolve all relationships with limited taxonomic density. As an example, relationships between the three major clades of rodents were not clearly resolved in a recent analysis based on high quality genomes (Foley *et al.* 2023).

To sum up, we think that targeted enrichment and sequencing of nuclear loci in general, and UCEs in particular, provide opportunities to resolve higher level relationships among geomyoids. In general, phylogenetic methods based on multispecies coalescent models appear particularly appropriate to analyze difficult nodes (*e. g.*, Kubatko and Degnan 2007; Liu and Edwards 2009), but challenges persist in the realm of multispecies coalescent methods (*e. g.* (Philippe *et al.* 2011; Meiklejohn *et al.* 2016). Recently, UCEs have been used to explore rapid radiations where short branches between speciation events are documented by few substitutions and may be obscured by extensive incomplete lineage sorting (Esselstyn *et al.* 2017; McLean *et al.* 2019; Parada *et al.* 2021).

Integration of molecular, morphological and paleontological data. It is perhaps not surprising that mitochondrial DNA data or taxonomically sparse nuclear data fail to resolve deep relationships with the Geomyoidea. Hafner *et al.* (2007: Figure 3) estimated that the divergence between Dipodomyinae, Perognathinae and Heteromyinae occurred in the Early Miocene and involved branching events some 20 to 22 MYA, and recent analyses of the fossil record of Geomyoidea are consistent with an Early Miocene, or even Late Oligocene divergence of the group (Samuels *et al.* 2023, and references therein). Short intervals between deep divergence events are typically difficult to recover with limited genetic sampling.

As summarized in the Introduction, proposals of heteromyid paraphyly relative to geomyids are not new but encompass three inter-related issues. First, paraphyly may result from the inclusion of stem fossil taxa in the Heteromyidae. For example, Korth (1994:182) noted that, "If *Proheteromys* is considered ancestral to the geomyids and is within the Heteromyidae, this also implies the derivation of the Geomyidae from a heteromyid ancestor."

Second, and restricting the focus on living geomyoids, crown Heteromyidae have at times been recovered as paraphyletic relative to Geomyidae. Early proposals of heteromyid paraphyly were based on qualitative assessments of dental traits (Wood 1931; Lindsay 1972). In general, as other characters were included, subsequent morphological analyses have supported reciprocal monophyly of crown heteromyids and geomyids (Wahlert 1985; Ryan 1989; Flynn *et al.* 2008), with the notable exception of an analysis of stapedial arteries (Brylski 1990).

Third, given the historical uncertainty in the placement of Geomyoidea among rodents, character polarity has been difficult to establish. Even recent phylogenetic reconstructions centered on fossil geomyoids (Samuels *et al.* 2023) do not include beavers as an outgroup. The sister-taxon relationship of Geomyoidea and Castoridae, which has been known for about two decades, has not been put to use to this end but is key to understanding the higher-level relationships of the Geomyoidea. It appears that the time is ripe to integrate the wealth of morphological information about living and fossil Geomyoidea with molecular phylogenetic data.

Ultimately, we want to infer phylogenetic relationships to help us understand how evolution has produced emblematic critters of North America, such as pocket mice, pocket gophers and kangaroo rats and mice, with their wealth of morphological, physiological, behavioral and ecological adaptations, building upon the remarkable insights provided by Jim Patton in his studies of these fascinating organisms.

Acknowledgments

Our understanding of the evolutionary biology of the Geomyoidea has been greatly impacted by Jim Patton's work throughout his remarkable career, beginning with the articles stemming from his graduate work (Patton 1967), and including the recent, high-quality genomes of *Perognathus longimembris* (Kozak *et al.* 2024) and *Thomomys bottae* (Voss *et al.* 2024), from which we extracted some of the data used in this study. We wish to express our admiration of Jim Patton for the breadth and impact of his scholarly contributions in mammalogy and evolutionary biology, for his broader contributions to the development of science internationally, and for his generosity as a mentor of innumerable students and scholars. In addition, Enrique Lessa wishes to express his personal gratitude to both Jim and Carol Patton for their encouragement and multi-faceted support throughout his career. We are grateful to Marjorie Matocq and an anonymous reviewer for their suggestions on earlier versions of this article.

Literature cited

Adkins, R. M., *et al.* 2001. Molecular phylogeny and divergence time estimates for major rodent groups: evidence from multiple genes. Molecular Biology and Evolution 18:777–791.

Adkins, R. M., A. H. Walton, and R. L. Honeycutt. 2003. Higher-level systematics of rodents and divergence time estimates based on two congruent nuclear genes. Molecular Phylogenetics and Evolution 26:409–420.

Bangs, M. R., and S. J. Steppan. 2022. A rodent anchored hybrid enrichment probe set for a range of phylogenetic utility: From order to species. Molecular Ecology Resources 22:1521–1528.

Bi, K., T. Linderoth, *et al.* 2013. Unlocking the vault: next-generation museum population genomics. Molecular Ecology 22:6018–6032.

Blanga-Kanfi, S., *et al.* 2009. Rodent phylogeny revised: analysis of six nuclear genes from all major rodent clades. BMC Evolutionary Biology 12:71–71.

Brylski, P. 1990. Development and evolution of the carotid circulation in geomyoid rodents in relationship to their craniomorphology. Journal of Morphology 204:33–45.

Camacho, C. *et al.* 2009. BLAST+: architecture and applications. BMC Bioinformatics 10:421.

Carleton, M. D., and G. G. Musser. 2005. Order Rodentia. Pp. 745–752, *in* Mammal species of the world: a taxonomic and geographic reference (Wilson, D. E., and D. M. Reeder, eds.). Third edit. Johns Hopkins University Press. Baltimore, U.S.A.

DeBry, R. W. 2003. Identifying conflicting signal in a multigene analysis reveals a highly resolved tree: the phylogeny of Rodentia (Mammalia). Systematic Biology 52:604–617.

DeBry, R. W., and R. M. Sagel. 2001. Phylogeny of Rodentia (Mammalia) inferred from the nuclear-encoded gene IRBP. Molecular Phylogenetics and Evolution 19:290–301.

D'Elía, G., P. H. Fabre, and E. P. Lessa. 2019. Rodent systematics in an age of discovery: recent advances and prospects. Journal of Mammalogy 100:852–871.

Esselstyn, J. A., *et al.* 2017. Investigating difficult nodes in the placental mammal tree with expanded taxon sampling and thousands of ultraconserved elements. Genome Biology and Evolution 9:2308–2321.

Fabre, P.-H., E. J. P. Douzery, and L. Hautier. 2015. A synopsis of the rodent systematics, biogeography and diversification history. Pp. 19–69, *in* Evolution of the rodents: advances in phylogeny, functional morphology and development (Cox, P. G., and L. Hautier, eds.). Cambridge University Press. Cambridge, United Kingdom.

Fabre, P.-H., *et al.* 2012. A glimpse on the pattern of rodent diversification : a phylogenetic approach. BMC Evolutionary Biology 12:88–88.

Faircloth, B. C. 2016. PHYLUCE is a software package for the analysis of conserved genomic loci. Bioinformatics 32:786–788.

Flynn, L. J., *et al.* 2019. Rodent Suborders. Fossil Imprint 75:292–298.

Flynn, L. J., E. H. Lindsay, and R. A. Martin. 2008. Geomorpha. Pp. 428–455, *in* Evolution of Tertiary Mammals of North America (Janis, C. M., G. F. Gunnell & M. D. E. Uhen, eds.). Cambridge University Press. Cambridge, United Kingdom.

Foley, N. M. *et al.* 2023. A genomic timescale for placental mammal evolution. Science 380:eabl8189.

Hafner, J. C. *et al.* 2007. Basal clades and molecular systematics of heteromyid rodents. Journal of Mammalogy 88:1129–1145.

Harder, A. M., *et al.* 2022. High-quality reference genome for an arid-adapted mammal, the banner-tailed kangaroo rat (*Dipodomys spectabilis*). Genome Biology and Evolution 14:evac005.

Huchon, D., F. M. Catzeflis, and E. J. Douzery. 1999. Molecular evolution of the nuclear von Willebrand factor gene in mammals and the phylogeny of rodents. Molecular Biology and Evolution 16:577–589.

Kalyaanamoorthy, S., *et al*. 2017. ModelFinder: fast model selection for accurate phylogenetic estimates. Nature Methods 14:587–589.

Katoh, K., and D. M. Standley. 2013. MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability. Molecular Biology and Evolution 30:772–780.

Korth, W. W. 1994. The Tertiary record of rodents in North America. Springer US, Boston, MA.

Kozak, K. M. *et al*. 2024. A highly contiguous genome assembly for the pocket mouse *Perognathus longimembris longimembris*. Journal of Heredity 115:130–138.

Kubatko, L. S., and J. H. Degnan. 2007. Inconsistency of phylogenetic estimates from concatenated data under coalescence. Systematic Biology 56:17–24.

Lindsay, E. H. 1972. Small mammal fossils from the Barstow Formation, California. University of California Publications in Geological Sciences 93:1–104.

Liu, L., and S. V. Edwards. 2009. Phylogenetic analysis in the anomaly zone. Systematic Biology 58:452–460.

Lok, S. *et al*. 2017. De novo genome and transcriptome assembly of the Canadian beaver (*Castor canadensis*). G3 (Bethesda, Md.) 7:755–773.

Mammal Diversity Database. 2023. Mammal Diversity Database (Version 1.12.1) [Data set]. doi:10.5281/zenodo.7830771.

McCormack, J. E., *et al*. 2012. Ultraconserved elements are novel phylogenomic markers that resolve placental mammal phylogeny when combined with species-tree analysis. Genome Research 22:746–754.

McLean, B. S., *et al*. 2019. Impacts of inference method and data set filtering on phylogenomic resolution in a rapid radiation of ground squirrels (Xerinae: Marmotini). Systematic Biology 68:298–316.

Meiklejohn, K. A., *et al*. 2016. Analysis of a rapid evolutionary radiation using ultraconserved elements: evidence for a bias in some multispecies coalescent methods. Systematic Biology 65:612–627.

Meng, G., Y. Li, C. Yang, and S. Liu. 2019. MitoZ: a toolkit for animal mitochondrial genome assembly, annotation and visualization. Nucleic Acids Research 47:e63.

Minh, B. Q. *et al*. 2020. IQ-TREE 2: New models and efficient methods for phylogenetic inference in the genomic era. Molecular Biology and Evolution 37:1530–1534.

Montgelard, C., *et al*. 2008. Suprafamilial relationships among Rodentia and the phylogenetic effect of removing fast-evolving nucleotides in mitochondrial, exon and intron fragments. BMC Evolutionary Biology 8:321.

Murphy, W. J. *et al*. 2001. Resolution of the early placental mammal radiation using Bayesian phylogenetics. Science 294:2348–2351.

Parada, A., J. Hanson, and G. D'Eiía. 2021 . Ultraconserved elements improve the resolution of difficult nodes within the rapid radiation of Neotropical sigmodontine rodents (Cricetidae: Sigmodontinae). Systematic Biology 70:1090–1100.

Patton, J. L. 1967. Chromosome studies of certain pocket mice, genus *Perognathus*. Journal of Mammalogy 48:27–37.

Peterson, B. K., *et al*. 2012. Double digest RADseq: an inexpensive method for de novo SNP discovery and genotyping in model and non-model species. Plos One 7:e37135–e37135.

Philippe, H. *et al*. 2011. Resolving difficult phylogenetic questions: why more sequences are not enough. PLoS biology 9:e1000602.

Ryan, J. M. 1989. Comparative myology and phylogenetic systematics of the Heteromyidae (Mammalia, Rodentia). Miscellaneous Publications of the Museum of Zoology, University of Michigan 176:1-103.

Samuels, J. X., J. J.-M. Calede, and R. M. Hunt, Jr. 2023. The earliest dipodomyine heteromyid in North America and the phylogenetic relationships of geomorph rodents. PeerJ 11:e14693.

Sayyari, E., and S. Mirarab. 2016. Fast coalescent-based computation of local branch support from quartet frequencies. Molecular Biology and Evolution 33:1654–1668.

Spradling, T. A., *et al*. 2004. DNA data support a rapid radiation of pocket gopher genera (Rodentia: Geomyidae). Journal of Mammalian Evolution 11:105–125.

Swanson, M. T., C. H. Oliveros, and J. A. Esselstyn. 2019. A phylogenomic rodent tree reveals the repeated evolution of masseter architectures. Proceedings of the Royal Society B: Biological Sciences 286:20190672.

Tomasco, I. H. *et al*. 2024. Underground speciation: unraveling the systematics and evolution of the highly diverse tuco-tucos (genus *Ctenomys*) with genomic data. Molecular Phylogenetics and Evolution 199:108163.

Upham, N. S., J. A. Esselstyn, and W. Jetz. 2019. Inferring the mammal tree: species-level sets of phylogenies for questions in ecology, evolution, and conservation. PLOS Biology 17:e3000494.

Veniaminova, N. A., N. S. Vassetzky, and D. A. Kramerov. 2007 . B1 SINEs in different rodent families 89:678–686.

Voss, E. R. *et al*. 2024. De novo genome assembly of a geomyid rodent, Botta's pocket gopher (*Thomomys bottae bottae)*. The Journal of Heredity esae045.

Wahlert, J. H. 1985. Skull morphology and relationships of geomyoid rodents. American Museum Novitates 2812:1–20.

Wilson, D. E., and D. M. Reeder (eds.). 2005. Mammal species of the world. A taxonomic and geographic reference. 3rd. edition. Johns Hopkins University Press, Baltimore.

Wood, A. E. 1931. Phylogeny of heteromyid rodents. American Museum Novitates 501:1–20.

Wood, A. E. 1955. Classification of rodents. Journal of Mammalogy 36:165–187.

Zhang, C., and S. Mirarab. 2022. Weighting by gene tree uncertainty improves accuracy of quartet-based species trees. Molecular Biology and Evolution 39:msac215.

Zhang, C., M. Rabiee, E. Sayyari, and S. Mirarab. 2018. ASTRAL-III: polynomial time species tree reconstruction from partially resolved gene trees. BMC Bioinformatics 19:153.

Associate editors: Marjorie Matocq and Eileen Lacey
Submitted: July 5, 2024; Reviewed: November 22, 2024
Accepted: December 16, 2024; Published on line: January 31, 2025

Supplementary materials

The datasets (two mitochondrial, 2 UCE), partition schemes, and additional details on the various loci, along with the results of phylogenetic analyses are available at https://doi.org/10.6084/m9.figshare.28046777.v1.