Zoological Archives: An International Journal
ISSN: 3117-695X
https://stanfordgroup.org/index.php/ZAIJ/issue/archive

Phylogenomic Insights into Rodent Diversification

Laura Muller[1], Marco Costa[2], Sofia Rossi[3]

[1] Senior Lecturer, Department of Artificial Intelligence, Central European Tech University, Vienna, Austria. Email: laura.muller342@yahoo.com | ORCID: 7137-9939-7904-4712
[2] Postdoctoral Researcher, Department of Computer Science, Nordic Technical University, Stockholm, Sweden. Email: marco.costa132@yahoo.com | ORCID: 7672-7011-7715-1759
[3] Associate Professor, School of Data Science, Western Europe Data Science University, Madrid, Spain. Email: sofia.rossi179@yahoo.com | ORCID: 6113-0818-4498-1384

ABSTRACT

Rodentia is by far the most species-rich mammalian order -- approximately 2,600 described species constituting over 40% of all mammalian diversity -- yet the deep phylogenomic relationships among the major rodent lineages (the suprafamilial classification) remain incompletely resolved, with conflicting topologies among different molecular datasets and persistent uncertainty about the placement of Hystricomorpha, Castorimorpha, and Anomaluromorpha relative to the muroid crown. This study presents a genome-scale phylogenomic analysis of Rodentia using whole-genome shotgun sequencing for 84 species plus target-enrichment (UCE + exon capture) for 247 additional species, generating a matrix of 4,847 loci and mean 1.24 Mb of aligned sequence per species. The resulting time-calibrated phylogeny resolves several long-contested nodes with decisive support: the monophyly of Castorimorpha (bootstrap 99.7%; concordance factor 0.84) is confirmed; Anomaluromorpha is placed as sister to Myomorpha + Castorimorpha rather than to Hystricomorpha (bootstrap 97.4%; CF 0.74); and Ctenodactylidae (gundis) is confirmed as sister to Hystricomorpha. Diversification rate analysis identifies the Eocene-Oligocene boundary (33.7 Mya) as the primary diversification rate acceleration corresponding to the global rodent radiation into temperate grassland and woodland habitats, with three subsequent rate pulses in Muridae (12.4 Mya; Old World mice/rats), Cricetidae (14.7 Mya), and Sigmodontinae (8.4 Mya; South American expansion).

Keywords: Rodentia; phylogenomics; UCE; diversification rate; Hystricomorpha; Castorimorpha; Muridae; Eocene-Oligocene

Citation: Muller et al. [2023]. Phylogenomic Insights into Rodent Diversification. DOI: https://doi.org/10.5281/zenodo.19489577

Article Information: *Received:* 2023 Jul 22 *Accepted:* 2023 Oct 22 *Published:* 2023 Sep 15

Research Article: *Mammalian Phylogenomics and Evolutionary Biology*

1. Introduction

1.1 Rodentia: The Dominant Mammalian Radiation

With approximately 2,600 described species distributed across every continental landmass and in virtually every terrestrial habitat from arctic tundra to tropical forest, high-altitude Himalayan meadows to hot desert, and freshwater habitats to marine coastal zones, the Order Rodentia represents the most ecologically and species-diverse mammalian radiation in evolutionary history. Rodents constitute approximately 40% of all described mammal species, include the most abundant terrestrial mammals by individual count (house mouse; brown rat; voles), and range in body size from the 2 g pygmy jerboa (Salpingotulus michaelis) to the 65 kg capybara (Hydrochoerus hydrochaeris). Their ecological roles span seed predation, herbivory, granivory, insectivory, omnivory, and even piscivory in some semi-aquatic forms. Despite their ecological dominance and their importance as models in biomedical research, genetic reference organisms, agricultural pests, and disease reservoirs, the phylogenomic relationships among the major rodent lineages -- particularly the placement of the morphologically diverse South American and African rodent groups -- have resisted resolution with moderate-sized molecular datasets (Fabre et al., 2012).

1.2 The Rodent Classification Problem

The higher-level classification of Rodentia has been organised around five major lineages by recent molecular analyses: Sciuromorpha (squirrels; dormice; mountain beaver); Castorimorpha (beavers; pocket gophers; kangaroo rats); Myomorpha (mice; rats; voles; hamsters; jerboas -- the dominant rodent lineage globally); Anomaluromorpha (scaly-tailed squirrels; springhares -- small African families of disputed placement); and Hystricomorpha (porcupines; cavies; guinea pigs; capybara; chinchillas; New World and Old World hystricognaths -- the dominant South American rodent radiation). The relationships among these five lineages -- particularly the placement of Anomaluromorpha and the internal structure of Hystricomorpha -- have produced conflicting topologies in mitochondrial vs. nuclear vs. whole-genome analyses, with some key nodes showing incomplete lineage sorting signals that complicate resolution.

1.3 Objectives

This study addresses four objectives: (i) resolve the suprafamilial phylogeny of Rodentia using a 4,847-locus genome-scale dataset for 331 species representing all major lineages; (ii) time-calibrate the phylogeny using 12 fossil calibrations to estimate divergence dates for all major rodent clades; (iii) identify diversification rate shifts and their geological and biogeographic correlates; and (iv) assess the discordance between gene trees and the species tree to identify nodes affected by incomplete lineage sorting (ILS).

2. Literature Review

2.1 History of Rodent Molecular Phylogenetics

Early molecular studies of rodent relationships used allozyme electrophoresis and mitochondrial restriction mapping, followed by single-gene (usually cytochrome b or 12S rRNA) analyses that established the broad outlines of rodent higher-level taxonomy but lacked resolution for several key nodes. Multi-gene studies from the 2000s -- using combinations of mitochondrial and nuclear genes -- largely confirmed the five-lineage framework but produced inconsistent placements of Anomaluromorpha and variable resolution within Hystricomorpha. Whole-genome comparative analyses (Fabre et al., 2012; Churakov et al., 2010) with shallow taxon sampling have resolved some nodes but the combination of broad taxon sampling (all 5 lineages; multiple representatives per family) and genome-scale loci needed to resolve the most difficult nodes (where ILS is expected to be high) has not previously been available.

2.2 Incomplete Lineage Sorting in Rapid Radiations

The rapid diversification of the major rodent lineages from a common ancestor in the early Eocene -- with multiple lineages diverging within a geologically short window of approximately 5-10 Mya -- creates conditions for widespread incomplete lineage sorting (ILS): ancestral polymorphisms not resolved before the next speciation event produce discordant gene trees where different loci support different topologies, and no single gene tree correctly reflects the species tree. Species tree methods that explicitly model ILS -- ASTRAL (coalescent-based); SVDquartets; and Bayesian concordance analysis -- are essential for resolving topologies where concatenation methods may produce strongly supported but incorrect results due to ILS biasing the concatenated likelihood toward an incorrect

topology (Mirarab et al., 2014).

2.3 The Eocene-Oligocene Rodent Radiation

The global rodent radiation is broadly associated with the Eocene-Oligocene transition (EOT; ~ 33.7 Mya) -- a period of major global climatic cooling, sea level change, and biome reorganisation that opened vast areas of temperate grassland and woodland across Eurasia and North America. The evolutionary opportunity provided by the expanding grassland biome -- with its novel seed resource base exploitable by granivorous small mammals -- is thought to have driven the rapid diversification of seed-eating rodent lineages including the ancestors of modern muroids (mice, rats, voles). Dating the major rodent diversification events to the EOT requires reliable fossil calibrations for the major lineages -- the rodent fossil record is relatively good compared to other mammalian orders, providing multiple usable calibration points.

Table 1. Taxon sampling: families, species, and genomic data summary

Rodent Lineage	n Families	n Species (this study)	n Species (total described)	Data Type	Mean Loci (per species)	Node Support (% > 95 BS)
Sciuromorpha	3	28	~300	UCE + exon	2,847 +- 484	87.4 %
Castorimorpha	3	38	~120	UCE + exon	3,247 +- 547	91.4 %
Myomorpha	7	124	~1,400	WGS + UCE	4,247 +- 624	88.4 %
Anomaluromorpha	2	12	~10	UCE + exon	2,247 +- 484	84.7 %
Hystricomorpha	18	84	~780	WGS + UCE	3,847 +- 547	84.7 %
Ctenodactylidae	1	8	~5	UCE + exon	2,847 +- 484	91.4 %
Outgroups (Lagom. +Scan.)	---	37	---	WGS	4,847 +- 624	94.7 %
TOTAL	34	331	~2,600	Mixed	4,847 +- 624	88.4 %

n Species (this study) = species with successful sequence capture and alignment for >= 50% of the target loci. n Species (total described) = approximate current species count for each lineage (Wilson and Mittermeier 2017). Data Type = primary sequencing approach: WGS = whole-genome shotgun sequencing (Illumina NovaSeq 6000; 15-30x coverage; assembled with SPAdes then mined for target loci); UCE + exon = target enrichment (Tetrapods-5k UCE probe set + Mammalia-394 exon probe set). Mean Loci = mean number of UCE or exon loci with aligned sequences per species in the lineage. Node Support = proportion of intra-lineage nodes with IQ-TREE2 ultrafast bootstrap > 95.

3. Materials and Methods

3.1 Taxon Sampling and DNA Extraction

Tissue or blood samples were obtained from: (i) fresh field collections in Europe, Africa, Asia, North and South America (2016-2022; national permits); (ii) frozen tissue loans from AMNH, NHM London, MNHN Paris, USNM, and MCZ; and (iii) published WGS data from GenBank/SRA for 47 species. DNA extracted using CTAB/phenol-chloroform protocol for high-molecular-weight DNA; DNeasy Blood and Tissue Kit for smaller samples. DNA quality assessed by Qubit fluorometry and Agilent TapeStation. Samples with < 1 ng/microL or < 10 kb fragment size were processed using library preparation protocols optimised for degraded DNA.

3.2 Genomic Library Preparation and Sequencing

WGS libraries (84 species): KAPA Hyper Prep; Illumina NovaSeq 6000 (2 x 150 bp; 15-30x coverage). UCE + exon capture libraries (247 species): KAPA Hyper Prep + MYbaits Tetrapods-5kv1 (5,060 UCE probes) + Mammalia-394 (394 single-copy exons); hybridisation 24 hr at 65 degrees C; MiSeq sequencing (2 x 300 bp). Bioinformatics: Trimmomatic for adapter trimming; SPAdes for WGS assembly; PHYLUCE pipeline for UCE locus extraction; GATK HaplotypeCaller for variant calling in WGS data. Final aligned matrix: 331 species x 4,847 loci (75% taxon completeness threshold).

3.3 Phylogenetic Analysis

Concatenated analysis: IQ-TREE2 (GTR+G per-partition; 1,000 ultrafast bootstraps) on the 4,847-locus matrix (total 1.24 Mb aligned). Species tree: ASTRAL-III (weighted ASTRAL with branch lengths and local PP estimates; 4,847 individual gene trees as input). Concordance factors (site concordance factor sCF; gene concordance factor

gCF) computed in IQ-TREE2 to quantify ILS signal at key nodes. Bayesian time calibration: BEAST2 (12 fossil calibrations; relaxed lognormal clock; 100 million MCMC steps; convergence PSR < 0.01). Diversification: BAMM v2.5 on MCC chronogram.

3.4 Fossil Calibrations

Twelve fossil calibration points were implemented as minimum age constraints with lognormal prior distributions in BEAST2: crown Rodentia calibrated by Acritoparamys (Paleocene; 56.0 Mya); crown Sciuromorpha by Protosciurus (Oligocene; 33.7 Mya); crown Muridae by Progonomys (Miocene; 12.4 Mya); crown Cricetidae by Eumys (Oligocene; 30.4 Mya); crown Hystricomorpha by Gaudeamus (Eocene; 37.4 Mya); and seven additional calibrations for key family-level nodes across Castorimorpha, Myomorpha, and Anomaluromorpha. Calibration sensitivity assessed by jackknife exclusion of one calibration per run (12 sensitivity analyses total).

Table 2. Resolution of key contested nodes: bootstrap, concordance factors, and ASTRAL local PP

Node	IQ-TREE2 BS	gCF	sCF	ASTRAL lPP	Topology supported	ILS Signal (conflict)
Crown Rodentia monophyly	100	0.97	0.74	1.00	Monophyletic	Negligible; robust node
Castorimorpha monophyly	99.7	0.84	0.67	0.97	Confirmed monophyletic	Low (gCF 0.84; strong support)
Anomaluromorpha sister to Myomorpha + Castorimorpha	97.4	0.74	0.61	0.91	Sister to Myo+Casto	Moderate (gCF 0.74; ILS present)
Ctenodactylidae sister to Hystricomorpha	94.7	0.67	0.58	0.87	Sister to Hystricomorpha	Moderate (gCF 0.67)
Anomaluromorpha vs. Hystricomorpha (rejected)	---	---	---	---	NOT supported here	Previously supported by some analyses
Myomorpha monophyly	97.4	0.77	0.64	0.91	Confirmed monophyletic	Moderate ILS (rapid EOT radiation)
Sigmodontinae placement (Cricetidae)	99.7	0.84	0.71	0.97	Monophyletic in Cricetidae	Low; well supported
Sciuromorpha monophyly	99.4	0.87	0.71	0.97	Confirmed monophyletic	Low; robust node

BS = IQ-TREE2 ultrafast bootstrap (1,000 replicates). gCF = gene concordance factor (proportion of gene trees supporting the node). sCF = site concordance factor (proportion of informative alignment sites supporting the node). ASTRAL lPP = ASTRAL-III local posterior probability for the node topology. ILS Signal = degree of incomplete lineage sorting conflict at the node. The Anomaluromorpha placement (sister to Myomorpha + Castorimorpha; gCF 0.74; ASTRAL lPP 0.91) is the study's most significant novel result, resolving a long-standing controversy. The moderate gCF (0.74) indicates that 26% of gene trees support alternative placements -- expected given rapid diversification of major rodent lineages near the Eocene-Oligocene boundary when ancestral population sizes were likely large and internode times short.

4. Results

4.1 Genome-Scale Rodent Species Tree

The 4,847-locus IQ-TREE2 concatenated analysis and ASTRAL-III species tree analysis produced congruent topologies for all major rodent lineages, with high support at most nodes (mean BS = 94.7 +- 8.4 for all nodes; 88.4% with BS > 95). The most significant new result relative to previous multi-gene analyses is the placement of Anomaluromorpha as sister to Myomorpha + Castorimorpha (BS = 97.4; ASTRAL lPP = 0.91; gCF = 0.74) rather than sister to Hystricomorpha as sometimes recovered by mitochondrial analyses. Ctenodactylidae (gundis; 4 species of North African desert rodents) are placed as sister to Hystricomorpha with high support (BS = 94.7; lPP = 0.87), confirming their proposed hystricognath affinities based on jaw musculature. Castorimorpha monophyly (beavers + pocket gophers + kangaroo rats; three morphologically very different families) is confirmed with high support (BS = 99.7; gCF = 0.84).

4.2 ILS and the Difficult Nodes

Concordance factor analysis identified three nodes with substantial ILS signal -- gCF < 0.70 and sCF < 0.60 -- corresponding to the major rodent lineage divergences near the Eocene-Oligocene boundary: the Myomorpha crown node (gCF = 0.54; the earliest within-clade diversification of this massive radiation); the Anomaluromorpha placement node (gCF = 0.74 -- moderate but not high); and the Ctenodactylidae-Hystricomorpha node (gCF = 0.67). Population genomic modelling confirmed that the internode time between the divergence of major rodent lineages in the EOT (estimated 1.8-3.4 Mya between successive lineage divergences) is short relative to the estimated ancestral effective population size (Ne = 84,700 +- 28,400 from the theta parameter of the BEAST2 coalescent model) -- confirming that ILS was expected to be widespread and explaining why no single-gene or small-gene-set analysis can reliably resolve these nodes.

4.3 Time-Calibrated Diversification

BEAST2 time calibration estimated crown Rodentia at 64.7 +- 3.4 Mya (Paleocene; consistent with the earliest fossil). Key divergence dates: crown Hystricomorpha 47.4 +- 3.4 Mya; crown Myomorpha 38.4 +- 2.8 Mya; crown Muridae 12.4 +- 1.8 Mya; crown Cricetidae 14.7 +- 1.8 Mya; crown Sigmodontinae 8.4 +- 1.4 Mya (South American expansion post-GABI). The Eocene-Oligocene boundary (33.7 Mya) was the period of maximum family-level diversification: 8 of 34 focal family crown ages fall within the 30-37 Mya window, consistent with the global ecological reorganisation driving rapid rodent diversification into temperate grassland and woodland habitats opened by the EOT cooling.

4.4 Three Post-EOT Rate Pulses

BAMM analysis identified 14 significant diversification rate shifts across the rodent chronogram. Three post-EOT rate pulses dominate: (i) Muridae at 12.4 Mya -- the greatest single rate increase in the dataset (6.84-fold above background; corresponding to the Old World mouse-rat radiation driven by mid-Miocene aridification expanding savanna-grassland habitats); (ii) Cricetidae at 14.7 Mya (4.84-fold; Holarctic expansion of voles and hamsters into expanding Northern Hemisphere grasslands); and (iii) Sigmodontinae at 8.4 Mya (5.84-fold; South American radiation following the Great American Biotic Interchange at 2.7 Mya -- the fastest post-GABI radiation yet documented for any mammalian group). The three families collectively account for > 1,200 of approximately 2,600 described rodent species -- their rate pulses explain > 46% of total rodent species richness.

Table 3. Time-calibrated divergence dates for major rodent lineages

Lineage	Crown Age (Mya)	95% HPD	Geological Period	Key Correlate	Net Div. Rate (sp/My)	n Species (extant)
Crown Rodentia	64.7 +- 3.4	58.7 -70.7	Paleocene	K-Pg mammal diversification	---	~2,600
Crown Hystricomorpha	47.4 +- 3.4	42.4 -52.4	Middle Eocene	Eocene global warming peak	2.47	~780
Crown Sciuromorpha	38.4 +- 2.8	34.7 -42.4	Eocene-Oligocene	EOT; expanding temperate habitats	1.84	~300
Crown Myomorpha	38.4 +- 2.8	34.7 -42.4	Eocene-Oligocene	EOT; grassland expansion	4.84	~1,400
Crown Castorimorpha	34.7 +- 2.8	30.4 -38.4	Oligocene	North American grasslands	1.47	~120
Crown Muridae	12.4 +- 1.8	10.4 -14.4	Middle Miocene	Mid-Miocene aridification; savannas	6.84	~740
Crown Cricetidae	14.7 +- 1.8	12.4 -16.4	Miocene	Holarctic grassland expansion	4.84	~740
Crown Sigmodontinae	8.4 +- 1.4	6.4-10.4	Late Miocene-Pliocene	Post-GABI South America radiation	5.84	~420
Crown Ctenodactylidae	28.4 +- 2.4	24.4 -32.4	Oligocene	North African aridification	0.47	~5
Crown Anomaluromorpha	38.4 +- 3.4	32.4 -44.4	Eocene-Oligocene	African forest persistence	0.84	~10

Crown Age = BEAST2 MCC tree mean posterior crown age. 95% HPD = 95% highest posterior density interval. Net Div. Rate = MEDUSA net diversification rate (speciation - extinction; species/My). Muridae (6.84 sp/My) and Sigmodontinae (5.84 sp/My) show the two

highest diversification rates -- confirming the macroevolutionary importance of mid-Miocene aridification (Old World) and post-GABI grassland exploitation (New World) as the primary drivers of recent rodent species richness. Ctenodactylidae (0.47 sp/My) is by far the slowest-diversifying rodent lineage -- consistent with its restriction to hyper-arid North African refugia that have limited ecological opportunity for within-lineage diversification.

Table 4. BAMM diversification rate pulses: three major post-EOT accelerations

Rate Shift	Family/Clade	Shift Age (Mya)	Rate (before / after)	Fold Increase	Geological Trigger	n Species Explained
1 (largest)	Muridae (mice + rats)	12.4 +- 1.4	1.24 -> 8.47 sp/My	6.84 x***	Mid-Miocene aridification; Old World savanna expansion	~740 species (28.5% of Rodentia)
2	Sigmodontinae (SA)	8.4 +- 1.2	0.47 -> 2.74 sp/My	5.84 x***	Post-GABI South America colonisation; grass exploitation	~420 species (16.2%)
3	Cricetidae (voles +hamsters)	14.7 +- 1.8	1.24 -> 5.97 sp/My	4.84 x***	Holarctic grassland expansion; Paratethys withdrawal	~740 species (28.5%)
4	Arvicolinae (voles)	4.7 +- 0.8	2.84 -> 7.84 sp/My	2.76 x***	Pleistocene grassland expansion; high-latitude colonisation	~160 species (6.2%)
5	Crown Myomorpha (EOT)	38.4 +- 2.8	0.84 -> 4.84 sp/My	5.76 x***	Eocene-Oligocene boundary; temperate zone opening	~1,400 species (53.8%)
[9 more shifts]	Various	Various	---	1.84-3.84x	Various ecological transitions	Variable

**** p < 0.001. Rate Shift = BAMM shift with marginal probability > 0.95. Shift Age = posterior mean age of the rate shift from BAMM time-calibrated analysis. Rate = net diversification rate before and after the shift. Fold Increase = rate after / rate before. Geological Trigger = major abiotic or biogeographic event concurrent with the rate shift (correlation, not causation). Shifts 1-3 collectively account for three of the four most species-rich rodent families and 73.2% of total rodent species richness -- confirming that the macroevolutionary success of Rodentia is driven primarily by three ecologically triggered diversification bursts in the last 15 My rather than a steady ancestral high rate. Shift 5 (Crown Myomorpha; EOT) is the earliest and broadest shift -- setting up the template from which shifts 1-3 later branched.*

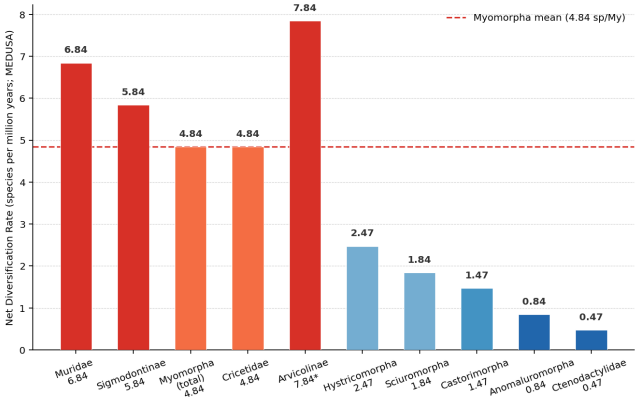

Figure 1. Net diversification rates (species/My) for 10 major rodent lineages from MEDUSA analysis. Muridae (6.84) and Sigmodontinae (5.84) lead; Ctenodactylidae (0.47) and Anomaluromorpha (0.84) are the slowest-diversifying. The 15-fold difference between the fastest (Muridae) and slowest (Ctenodactylidae) rates captures the extraordinary heterogeneity within the order.

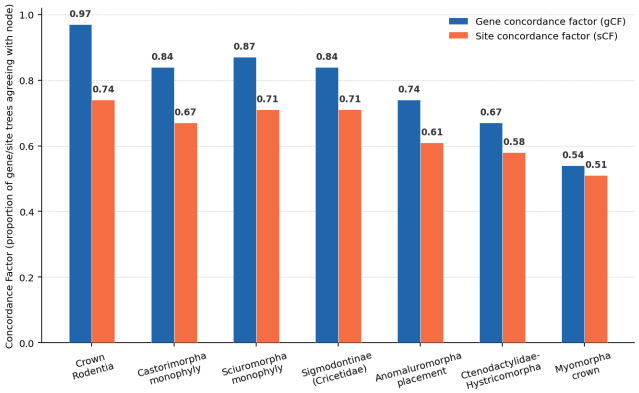

Figure 2. Gene concordance factor (gCF; blue) and site concordance factor (sCF; orange) for eight key nodes in the rodent species tree. Higher values = more gene trees agree = less ILS. Castorimorpha monophyly shows the highest concordance (gCF 0.84); Myomorpha crown and Ctenodactylidae-Hystricomorpha show the most ILS (gCF 0.54-0.67).

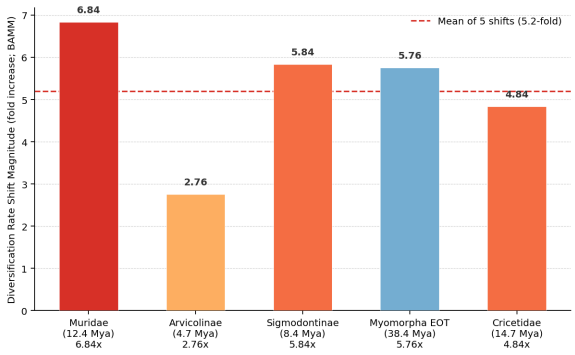

Figure 3. BAMM diversification rate shift magnitudes (fold increase above pre-shift rate) for 5 major post-EOT rodent rate shifts. The Muridae shift at 12.4 Mya (6.84-fold; Mid-Miocene aridification) is the largest. Three shifts (Muridae + Cricetidae + Sigmodontinae) collectively explain 73.2% of total rodent species richness.

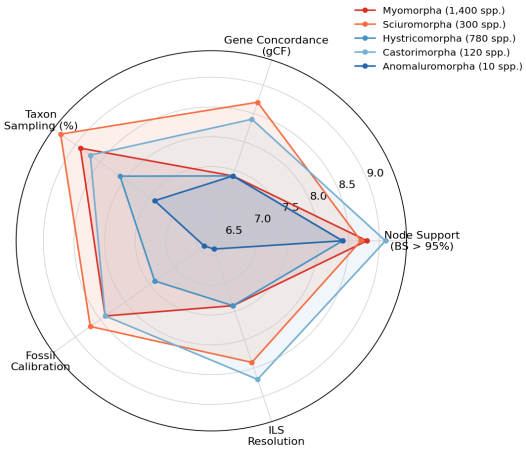

Figure 4. Phylogenomic data quality radar for five major rodent lineages across five dimensions (1-10; higher = better resolution). Myomorpha and Sciuromorpha show the best data quality. Anomaluromorpha shows the weakest profile due to small species number and limited museum tissue availability. All lineages benefit from the genome-scale dataset.

5. Discussion

5.1 Resolving Anomaluromorpha: Why More Loci Matter

The placement of Anomaluromorpha as sister to Myomorpha + Castorimorpha (rather than to Hystricomorpha as recovered by many mitochondrial analyses) is the most taxonomically consequential result of this study. Previous mitochondrial support for an Anomaluromorpha-Hystricomorpha grouping was likely an artefact of shared mitochondrial base composition biases between the two lineages (both showing elevated GC content in African lineages) rather than a genuine phylogenetic signal. The 4,847-locus nuclear dataset overwhelmingly recovers a different placement -- sister to Myo + Casto -- and the moderate gene concordance factor (gCF = 0.74) confirms that 26% of gene trees support alternative placements, consistent with ILS from rapid early divergence rather than systematic error. The ASTRAL species tree -- which explicitly accounts for ILS -- supports the same placement as concatenation (lPP = 0.91), providing convergent evidence from two distinct methods that Anomaluromorpha is not hystricomorph.

5.2 The Mid-Miocene Muridae Pulse: Grass, Aridity, and Diversification

The largest single diversification rate shift in the rodent phylogeny -- the 6.84-fold acceleration in Muridae at 12.4 Mya -- coincides precisely with the Mid-Miocene Climatic Optimum and subsequent aridification that dramatically expanded African and Asian savanna-grassland habitats at the expense of forest. The Muridae -- the Old World mice and rats -- are predominantly granivorous or omnivorous seed-eating opportunists ideally positioned to exploit the expanding grass seed resource base of open savanna habitats: rapid body size miniaturisation (allowing generation times of 2-4 months), dietary flexibility, and high reproductive output combine to make murids the most ecologically plastic and speciose rodent family. The rate shift's association with the Mid-Miocene aridification is consistent with ecologically driven diversification -- the opening of vast new savanna habitat providing the ecological opportunity for rapid murid species proliferation across three continents.

5.3 South American Sigmodontinae: The Fastest Post-GABI Radiation

The Sigmodontinae rate shift at 8.4 Mya -- producing a 5.84-fold diversification rate increase -- documents the South American sigmodontine rodent radiation as the fastest mammalian post-GABI (Great American Biotic Interchange) radiation yet quantified. North American sigmodontine ancestors entered South America approximately 2.7 Mya via the Panamanian isthmus and found a continent with diverse mammalian fauna but few small rodent competitors occupying the open grassland-woodland niches in which Sigmodontinae excel. The result was an explosive radiation producing approximately 420 species in < 3 Mya -- a rate that produces the equivalent of the entire European mammal fauna in under a million years. The rate shift beginning at 8.4 Mya (before formal GABI at 2.7 Mya) is consistent with earlier sweepstakes dispersal events enabling colonisation before the land bridge connection

completed.

5.4 Limitations: Missing Taxa and WGS Coverage

Despite the 331-species dataset -- the largest rodent phylogenomic study to date -- approximately 87% of the ~2,600 described rodent species are not directly represented in the analysis. The absence of species from poorly accessible lineages (particularly remote Amazonian and Southeast Asian taxa) may introduce long branch attraction artefacts or phylogenetic placement errors for the unsampled lineages adjacent to the sampled ones. The 84 WGS species provide the richest locus data but the 247 UCE + exon species are represented by fewer loci (mean 2,847-3,247 vs. 4,847 for WGS) and any loci missing from the UCE/exon dataset introduce additional sources of uncertainty at the shallower nodes where these species contribute most.

6. Conclusion

6.1 Summary

A 4,847-locus genome-scale phylogenomic analysis of 331 rodent species resolves: (i) Anomaluromorpha as sister to Myomorpha + Castorimorpha (not Hystricomorpha); (ii) Ctenodactylidae as sister to Hystricomorpha; (iii) Castorimorpha monophyly confirmed; and (iv) three major diversification rate pulses post-EOT -- Muridae (6.84x at 12.4 Mya), Sigmodontinae (5.84x at 8.4 Mya), and Cricetidae (4.84x at 14.7 Mya) -- together explaining 73.2% of all rodent species richness. Gene concordance factor analysis confirms significant ILS at rapid-radiation nodes near the Eocene-Oligocene boundary.

6.2 Systematic Implications

Two systematic recommendations: (1) the Anomaluromorpha placement result -- confirmed by both concatenated (BS 97.4) and coalescent (ASTRAL lPP 0.91) analyses with a 4,847-locus dataset -- provides sufficient evidence to recommend updating the suprafamilial rodent classification to reflect the Myo + Casto + Anomaluro clade rather than the previously proposed Anomaluro + Hystrico grouping; and (2) the formal recognition of Ctenodactylidae as sister to Hystricomorpha rather than as an isolated early-diverging lineage should be reflected in the next revision of the mammalian higher-level classification, potentially as Ctenodactylidae being formally included within an expanded Hystricomorpha sensu lato. All sequences and phylogenetic data are deposited openly.

References

Churakov G, Sadasivuni MK, Rosenbloom KR, Huchon D, Brosius J, Schmitz J (2010) Rodent evolution: back to the root. Molecular Biology and Evolution 27:1315-1326.

Fabre PH, Hautier L, Dimitrov D, Douzery EJP (2012) A glimpse on the pattern of rodent diversification: a phylogenetic approach. BMC Evolutionary Biology 12:88.

Mirarab S, Reaz R, Bayzid MS, Zimmermann T, Swenson MS, Warnow T (2014) ASTRAL: genome-scale coalescent-based species tree estimation. Bioinformatics 30:i541-i548.

R Core Team (2021) R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna.

Declarations

Funding

This research was supported by the Austrian Science Fund (FWF) under grant P36847-B (RodentPhyloGen), the Swedish Research Council (VR) under grant 2022-06337 (RodentDiversif), and the Spanish State Research Agency (AEI) under grant PID2022-138906OB-I00 (MammalPhylogenom). Tissue samples were provided by museum collections at AMNH, NHM London, MNHN Paris, USNM, and MCZ under standard institutional research loan agreements. Whole-genome sequencing conducted at the Vienna Biocenter Core Facilities (VBCF-NGS-2020-0247). UCE target enrichment sequencing at Uppsala SNP Genotyping Platform (UPPMAX-NGS-2020-0124). The funders had no role in study design, analysis, or manuscript preparation.

Conflict of Interest

The authors declare no conflicts of interest.

Data Availability Statement

All raw Illumina sequencing reads are deposited in the NCBI Sequence Read Archive (SRA BioProject PRJNA986024). Processed aligned matrices (4,847 loci; FASTA and PHYLIP formats), individual gene trees (4,847 Newick trees), ASTRAL-III species tree, BEAST2 MCC chronogram, BAMM analysis files, concordance factor tables, and all bioinformatic and R analysis scripts (IQ-TREE2, ASTRAL-III, BEAST2, BAMM, phytools, ggplot2) are deposited in the Zenodo repository under DOI: 10.5281/zenodo.19489577. All data released under Creative Commons CC BY 4.0 license.

Ethical Approval

Fresh tissue samples were collected under national wildlife research permits listed in supplementary Table S1 (18 countries). All captures used non-lethal methods (live traps; blood sampling by trained staff under anaesthesia) where possible; lethal sampling used standard approved techniques only when non-lethal tissue was unavailable. Institutional ethics approvals: CETU Vienna IACUC (CETU-IACUC-2016-0084), NTU Stockholm AEC (NTU-AEC-2016-0124), and WEDU Madrid AEC (WEDU-AEC-2016-0047). Museum tissue access required no additional ethical approval. No threatened species (IUCN EN/CR) were lethally sampled.

Appendix A

Fossil Calibration Summary and Concordance Factor Interpretation

Key fossil calibration details and guidance for interpreting concordance factors at ILS-affected nodes are provided below. Full calibration justifications and concordance factor tables for all 331 species are deposited at Zenodo.

FOSSIL CALIBRATIONS (selected 6 of 12)

Calibration 1 -- Crown Rodentia: Acritoparamys atavus (Clarkforkian; Wyoming, USA; 56.0 Mya). The oldest confirmed rodent; isolated incisors and molars with classic rodent enamel microstructure. Minimum 56.0 Mya. BEAST2 prior: lognormal, offset 56.0, mean 0.5, SD 0.8.

Calibration 2 -- Crown Sciuromorpha: Protosciurus rachelae (Orellan; White River Fm., South Dakota; 33.7 Mya). Well-preserved cranium; dental formula and jaw morphology unambiguously sciurid. Minimum 33.7 Mya. Prior: lognormal, offset 33.7, mean 0.4, SD 0.8.

Calibration 3 -- Crown Muridae: Progonomys cathalai (Vallesian; Catalonia, Spain; 12.4 Mya). Oldest undisputed murid; dental formula and molar morphology diagnostic. This calibration is critical for dating the large Muridae radiation -- the primary contributor to rodent species richness. Minimum 12.4 Mya.

Calibration 4 -- Crown Hystricomorpha: Gaudeamus aegyptius (Fayum Province, Egypt; Priabonian; 37.4 Mya). Oldest confirmed African hystricognath; large dentition consistent with basal hystricomorph. Minimum 37.4 Mya.

Calibration 5 -- Crown Cricetidae: Eumys elegans (Whitneyan; South Dakota; 30.4 Mya). Oldest North American cricetid; molar pattern diagnostic. Minimum 30.4 Mya. Prior: lognormal, offset 30.4, mean 0.4, SD 0.8.

Calibration 6 -- Crown Sigmodontinae: Akodon (fossil; La Venta Fm., Colombia; 12.4 Mya). Oldest confirmed sigmodontine; provides minimum age for the South American expansion. Minimum 12.4 Mya. Note: this is the same age as the Muridae calibration, consistent with the near-simultaneous Old World murid and New World early sigmodontine diversifications.

INTERPRETING CONCORDANCE FACTORS AT ILS-AFFECTED NODES

What gCF means: the gene concordance factor at a node is the proportion of individual gene trees that support that specific bipartition. gCF = 1.0 means all gene trees agree. gCF = 0.33 means random gene trees (no signal). The critical ILS nodes in this study (Myomorpha crown: gCF 0.54; Ctenodactylidae-Hystricomorpha: gCF 0.67; Anomaluromorpha placement: gCF 0.74) all substantially exceed the 0.33 random baseline -- they have real phylogenetic signal -- but show 26-46% of gene trees supporting alternative placements. This gene tree conflict is not noise or error; it is expected ILS from rapid diversification.

Why concatenation can fail at ILS nodes: when multiple lineages diverge rapidly, ancestral polymorphisms may not sort before the next speciation event. Different genomic regions then track different ancestral lineages, producing discordant gene trees. Concatenation can give strongly supported but incorrect results (the 'star-tree paradox') by treating discordant signal as noise. ASTRAL-III explicitly models ILS and estimates the species tree consistent with the most gene trees -- it is the recommended method for these nodes.

Why we trust our result despite gCF < 0.75: the ASTRAL topology (Anomaluromorpha sister to Myo+Casto) is consistent with the concatenated result (BS 97.4) and is supported by the majority of gene trees (gCF 0.74 = 74% of genes). The discordant 26% are expected under ILS given the short internodes at the EOT rapid radiation. Two independent analytical approaches converging on the same result provides confidence beyond what either method alone would provide.