Digging for the spiny rat and hutia phylogeny using a gene capture approach, with the description of a new mammal subfamily

Maxime Courcelle, Marie-Ka Tilak, Yuri L. R. Leite, Emmanuel J.P Douzery, Pierre-Henri Fabre

Title: **Digging for the spiny rat and hutia phylogeny using a gene capture approach, with the description of a new mammal subfamily**

Maxime Courcelle[a*], Marie-ka Tilak[a], Yuri L.R. Leite[b], Emmanuel J. P. Douzery[a] and Pierre-Henri Fabre[a],

[a] *ISEM, Univ Montpellier, CNRS, IRD, EPHE, Montpellier, France*

[b] *Universidade Federal do Espírito Santo, Vitória, Brasil*

* Corresponding author.

Email address: maxime.courcelle@umontpellier.fr

ABSTRACT

Next generation sequencing (NGS) and genomic database mining allow biologists to gather and select large molecular datasets well suited to address phylogenomics and molecular evolution questions. Here we applied this approach to a mammal family, the Echimyidae, for which generic relationships have been difficult to recover and often referred to as a star phylogeny. These South-American spiny rats represent a family of caviomorph rodents exhibiting a striking diversity of species and life history traits. Using a NGS exon capture protocol, we isolated and sequenced ca. 500 nuclear DNA exons for 35 species belonging to all major echimyid and capromyid clades. Exons were carefully selected to encompass as much diversity as possible in terms of rate of evolution, heterogeneity in the distribution of site-variation and nucleotide composition. Supermatrix inferences and coalescence-based approaches were subsequently applied to infer this family's phylogeny. The inferred topologies were the same for both approaches, and support was maximal for each node, entirely resolving the ambiguous relationships of previous analyses. Fast-evolving nuclear exons tended to yield more reliable phylogenies, as slower-evolving sequences were not informative enough to disentangle the short branches of the Echimyidae radiation. Based on this resolved phylogeny and on molecular and morphological evidence, we confirm the rank of the Caribbean hutias – formerly placed in the Capromyidae family – as Capromyinae, a clade nested within Echimyidae. We also name and define Carterodontinae, a new subfamily of Echimyidae, comprising the extant monotypic genus *Carterodon* from Brazil, which is the closest living relative of West Indies Capromyinae.

Keywords: Rodents; Echimyidae; Capromyinae; Exon capture; Phylogenomics;

1. INTRODUCTION

Over the past years, the advent of next-generation DNA sequencing (NGS) shifted the way biologists access and analyze molecular data (Zhang et al. 2011; Goodwin et al. 2016). Much emphasis has been put on the vast amount of data NGS can provide in a broad range of organisms. NGS are now routinely used in diverse biological fields for metabarcoding (Bohmann et al. 2014), population genomics (Yi et al. 2010), ancient DNA studies (Rohland et al. 2007) and phylogeny (Misof et al. 2014) purposes. Thus, NGS offers the possibility to choose molecular markers with desirable properties, rather than being limited by commonly used markers. This approach allows biologists to use the most appropriate molecular markers for their research, be it single nucleotide

polymorphism, ultra-conserved elements, or genes (Peterson et al. 2012; Faircloth et al. 2012; Yang et al. 2013). Among NGS-based techniques, sequence capture uses baits built from genomic references to specifically enrich sequencing libraries in genomic regions of interest. In the field of phylogenomics, sequence capture has been used to reliably gather datasets of hundreds to thousands of orthologous targets among various clades, such as mammals (Schweizer et al. 2016), birds (Smith et al. 2014), squamates (Bragg et al. 2018), lissamphibians (McCartney-Melstad et al. 2016; Newman and Austin 2016) and gymnosperms (Suren et al. 2016). Several methodological benefits of sequence capture can explain this recent interest in phylogenomics. Firstly, datasets as large as whole exomes can be acquired relatively cheaply (Bi et al. 2012). Secondly, capture performs well on degraded material, which enable the use of DNA from museum specimens (Fabre et al. 2014; McCormack et al. 2016). Thirdly, the technique tolerates a certain amount of nucleotide variation between baits and captured sequences, usually 15% of molecular divergence between a target and its reference (Bi et al. 2012). Thus, sequences from non-model or even extinct species might be captured if close relatives are available in genomic databases (Christmas et al. 2016; Carpenter et al. 2013; Enk et al. 2014).

Sequence capture offers the possibility to choose a dataset that fits the question being studied. Indeed, several characteristics linked to molecular evolution parameters have indeed proven useful to resolve difficult nodes of the phylogeny of metazoan organisms (Rodríguez-Ezpeleta et al. 2007; Philippe et al. 2011; Romiguier et al. 2013). Firstly, the evolutionary rate of molecular markers is defined by the number of substitutions that DNA, RNA and protein sequences accumulates through time. It thus impacts saturation and more generally homoplasy rates (Philippe and Laurent 1998). As such, a number of authors recommended to only use slower-evolving markers in phylogenetic studies (Steppan et al. 2004; Betancur-R et al. 2013; Nosenko et al. 2013). However, such slower-evolving genes may not accumulate enough substitutions to be phylogenetically informative (Townsend et al. 2012). Substitutions in conserved genes may also be under selective constraints themselves and, as a result, be more prone to homoplasy (Källersjö et al. 1999). This led many recent studies to combine both faster- and slower-evolving genes in order to benefit from their resolution at different taxonomic scales (de Knijff 2000; Delsuc et al. 2018). Secondly, distribution of substitutions across a gene may also have an impact on the number of homoplasies. If a site is more frequently subject to multiple hits, it is more likely to reach saturation, and to undergo convergences and/or reversals. Ideally, the variability of a phylogenetic marker should then be distributed equally among sites (Townsend and Naylor 2007). An example of such a

marker is the BReast CAncer 1 (BRCA1) gene, for which all codon positions were shown to accumulate substitutions at a similar rate and thus was a useful marker in mammal phylogeny (Delsuc et al. 2002; Meredith et al. 2011). Thirdly, the nucleotide composition of sequences has been linked to their potential to infer the correct phylogeny (Romiguier and Roux 2017). G+C-rich regions have for example been correlated to recombination through biased gene conversion (Galtier et al. 2001). By reducing the available substitution pool, biased gene conversion mechanically increases the rate of homoplasy (Romiguier and Roux, 2017). Moreover, recombination hotspots are regions where ancestral polymorphism is more often artifactually sorted, which leads to conflicts between gene trees and species tree (Hobolth et al. 2011). Some authors thus recommend favoring A+T-rich genes as phylogenetic markers (Betancur-R et al. 2013; Romiguier et al. 2013). As such, the lack of diversity of markers in phylogenetic analyses may hamper or at least affect inferences.

In recent phylogenomic studies, sequence capture has primarily targeted ultra-conserved elements (UCE) (Smith et al. 2014; Manthey et al. 2016; McCormack et al. 2016; Newman and Austin 2016; Branstetter et al. 2017). These slow-evolving regions are mostly conserved between species, making them an ideal target for sequence-capture baits. The phylogenetic signal carried by UCEs and their flanking regions has also been shown to be informative (Giarla and Esselstyn 2015; Gilbert et al. 2015). However, studies at lower taxonomic levels depend almost entirely on UCE flanking regions in order to gather enough variability between closely related species (Smith et al. 2014). Unlike UCEs (Bejerano et al. 2004; Katzman et al. 2007), nuclear exons cover a wider range of molecular evolution properties (Ranwez et al. 2007), while being conserved enough for a sequence capture kit to be designed (Bragg et al. 2016; Hugall et al. 2016). Thus, nuclear exons seem well-suited molecular markers to resolve difficult phylogenetic questions.

The generic relationships among the South-American spiny rats (Echimyidae), as well as their affinities with Caribbean hutias (Capromyidae) present such a challenging phylogenetic problem. Indeed, the phylogenetic relationship of hutias from the Greater Antilles have been widely debated and remain a major topic in rodent systematics (Woods and Howland 1979; McKenna and Bell 1997; Woods et al. 2001; Fabre et al. 2014; 2017). Earliest classifications of these rodent families, using molecular or morphological data, were incongruent and mostly highlighted their very fast early diversification (Lara et al. 1996; Leite and Patton 2002; Carvalho and Salles 2004; Emmons 2005). Capromyidae have then been classified either as a family or as a subfamily, considering both morphological (Woods and Howland 1979; McKenna and Bell 1997; Verzi et al.

2013) and molecular inferences (Upham and Patterson 2015; Fabre et al. 2014; 2017). Disentangling the relationships of hutias is therefore crucial not only to define their higher-taxonomic status but also to shed light on their potential biogeographical origin. Recent molecular phylogenies have suggested close affinities with the echimyid subfamily Euryzygomatomyinae (*Clyomys*, *Euryzygomatomys*, *Trinomys*; Fabre et al. 2017) from Eastern Brazil. Molecular dating analyses have supported an Early to Mid-Miocene divergence of hutias (Fabre et al. 2014; 2017) with a suspected dispersal from an Amazonian ancestor (Upham and Patterson 2012). However, the recent sampling of a fast-evolving echimyid from Eastern Brazil (*Carterodon sulcidens*) has hindered our view of their basal relationships (Upham and Patterson 2015; Fabre et al. 2017).

After more than two decades spent on this complex clade, the higher-level phylogeny of Echimyidae still appears unresolved with a basal polytomy between Echimyidae subfamilies and Capromyidae, with several nodes poorly supported (Fabre et al. 2017). On the one hand, recent morphological studies managed to gather large datasets of cranio-dental, jaws and external characters (Candela and Rasia 2012; Verzi et al. 2013), but those were shown to be very susceptible to homoplasy (Emmons 2005; Candela and Rasia 2012). Molecular studies on the other hand mainly relied on few nuclear and mitochondrial genes (Huchon et al. 2002; Galewski et al. 2005; Fabre et al. 2013; 2014; 2017; Álvarez et al. 2017), along with a mitogenomic study (Fabre et al. 2017). However, the haploidy and maternal inheritance of mitochondrial DNA reduce its effective population size, leading to a fast fixation of mutations (Brown et al. 1979), which increases the odds of saturation and makes sequences less reliable to infer deep relationships (Springer et al. 2001). Moreover, the absence of recombination in the mitochondrial genome means that all mitochondrial genes belong to a single loci. This is problematic as it is known that gene trees may differ from the species tree (Maddison and Wiens 1997; Huang et al. 2017). These points advocate the use of several diverse nuclear DNA loci as new molecular markers to explore the Echimyidae phylogeny.

In this paper we screened the OrthoMaM public database (Ranwez et al. 2007; Douzery et al. 2014) in order to gather a dataset of approximately 500 orthologous nuclear exons with contrasting molecular evolution properties in regard to their evolutionary rate, distribution of substitutions, and nucleotide content. This strategy would allow bypassing phylogenetic inference limitations induced by the analysis of a restricted class of marker (*e.g.* slow-evolving or G+C-rich markers). We then used an exon capture NGS protocol to amplify and sequence these coding genes markers for 35 non-model Echimyidae species, bringing together the largest molecular dataset to date for a rodent family. Bayesian and maximum-likelihood phylogenomics allowed us to generate

the first fully-resolved higher-level Echimyidae phylogeny. We therefore explored three questions regarding Echimyidae relationships: (i) How do we resolve the deepest divergences among echimyid subfamilies and other previously unresolved nodes of their phylogeny? (ii) Can we settle the taxonomic status of the Capromyidae (hutias)? and (iii) How do the evolutionary rate, substitution distribution and nucleotide content of nuclear DNA exons impact phylogenetic inferences?

2. MATERIALS AND METHODS

2.1. Taxonomic sampling

Biological samples were obtained from both museum and fresh tissue collections from the Museum of Comparative Zoology (MCZ: Cambridge, MA), the Museum of Vertebrate Zoology, University of California (MVZ: Berkeley, CA), the U.S. National Museum of Natural History (USNMH-Smithsonian: Washington DC), the Nationaal Natuurhistorisch Museum (Naturalis: Leiden, The Netherlands), the Royal Ontario Museum (ROM: Toronto, Canada), the Universidade Federal do Espírito Santo (UFES-CTA Animal Tissue Collection : Vitória, Brazil), and the University of Montpellier (UM-ISEM: Montpellier, FR), and are the same as those used in Fabre et al. (2017) for mitogenomics analyses. We re-identified the specimen of *Mesomys stimulax* RMNHN:MAM 21728 from Surinam, that was sequenced in Fabre et al. (2017). This specime was identified by Husson (1978; p. 438-440) and sampled near the Vier Gebroeders mountain near Sipaliwini in the South Eastern part of Nickerie District. This specimen was misidentified in Fabre et al. (2017) as *Mesomys stimulax*, and we re-identified it as *Mesomys hispidus*. In total, we obtained tissues for 35 taxa spanning almost all genera of Echimyidae and Capromyidae. In addition, we searched NCBI for complete genomes of two Phiomorpha rodents (*Heterocephalus glaber* and *Fukomys damarensis,* Genbank assembly accession GCA_000247695.1 and GCA_000743615.1 respectively*)*, and three Caviomorpha rodents *Chinchilla lanigera* (Chinchilloidea), *Cavia porcellus* (Cavioidea), and *Octodon degus* (Octodontidae), with accession numbers GCA_000276665.1, GCA_000151735.1 and GCA_000260255.1 respectively. These lineages span most Hystricognathi clades (Upham and Patterson 2015) and served as outgroups in our phylogeny analyses. Due to lack of DNA or absence of tissues, five other known Echimyinae genera (*Diplomys*, *Leiuromys*, *Lonchothrix, Santamartamys* and *Toromys*) were not included in this study.

2.2. Exon sampling

Prior to the sequence capture, we identified three molecular evolution characteristics that may influence the quality and quantity of phylogenetic signal of each nuclear DNA exon: its evolutionary rate, its among-site substitution rate heterogeneity and its nucleotide composition (Betancur-R et al. 2013; Townsend and Naylor 2007; Romiguier et al. 2013). The aim of this study was to capture 500 exons combining various parameter values of these three variables that are known to impact phylogenetic analyses. To do so, we used the orthologous gene database OrthoMaM (Douzery et al. 2014), which provides sequences and molecular evolution characteristics for mammalian nuclear exons. First, evolutionary rates of the exons at the hystricognath taxonomic scale were approximated by their Relative Rate of Evolution (RER) at the mammalian scale, as computed with a Super Distance Matrix approach and provided in OrthoMaM (Ranwez et al. 2007). Second, we used the shape parameter of the Gamma distribution that models the among-site heterogeneity in rates of evolution (α) as a proxy of variability distribution. When $0 < \alpha < 1$, the distribution is skewed towards a majority of invariable sites and a small number of positions accumulating most of the variability. In contrast when $\alpha > 1$, substitutions are more evenly distributed among sites (Yang 1996). Finally, nucleotide composition was estimated by the mean G+C content on the third codon positions of alignments (GC3) (Clay and Bernardi 2011). Two focal species were chosen as a starting point for this exon selection: *Mus musculus* is the most represented rodent in the OrthoMaM database, and *Cavia porcellus* is a caviomorph lineage more closely-related to the Echimyidae ingroup. Exons longer than 400 base pairs (bp) were classified in categories depending on these three molecular evolution parameters: four RER categories ranging from slower- to faster-evolving markers (i.e., $RER < 1$, $1 \leq RER < 2$, $2 \leq RER < 3$ and $3 \leq RER < 5$), four α categories ranging from strong to moderate among-site rate heterogeneity (i.e., $0.25 < \alpha < 0.50$, $0.50 \leq \alpha < 0.75$, $0.75 \leq \alpha < 1$, and $1 \leq \alpha$), as well as 61 GC3 categories ranging from 29% to 94% by steps of 1%. Among the 4 (RER) x 4 (α) x 61 (GC3) = 976 possible combinations, 483 exons matched the 3-parameter criteria and were finally identified and selected (see Table S1).

2.3. Exon capture and assembly

The hystricognath rodent *Octodon degus* is the closest relative of Echimyidae for which full genome data are available (Upham and Patterson 2015). We thus used this species' genome to design capture probes for our Echimyidae dataset. The sequences of the 483 target exons were extracted from *O. degus* and other outgroups genome with the TBLASTN software (Camacho et al. 2009) using amino-acid sequence of the *Mus musculus* and *Cavia porcellus* exons from OrthoMaM. The parameters used (wordsize = 3 and evalue < e^{-10}) allowed to recover slightly divergent

sequences, and only the best hit was retained for analysis. We then designed a capture probe set matching the 483 *O. degus* exons with a tiling (probe coverage) of 3x, which comprised 19,017 120-nucleotide long baits. DNA libraries for 35 Echimyidae species were subsequently prepared according to Tilak et al. (2014). Exons were then captured using a MyBaits (MYcroarray, Ann Arbour, USA) kit and following manufacturer protocol. Captured DNA was sequenced on a Illumina HiSeq 2500 platform (MGX, Montpellier, France). Raw 125-nucleotide long sequencing reads were paired in the Geneious software (Kearse et al. 2012). We used a Spearman rank correlation and an Analysis of Variance as implemented in the R software (R. Core Team 2018) to check if the age or location of tissues could explain the number of reads obtained. Adapters sequences were also trimmed and remaining reads were subsequently mapped onto *O. degus* reference sequences, while allowing for a maximum of 15% mismatch per read, and a consensus was produced. A second assembly run was performed using this consensus as reference while allowing only 5% mismatch in order to filter out spurious reads. After removal of duplicated reads, we excluded regions with coverage less than 3 from subsequent analyses.

2.4. Alignment and orthology assessment

Homologous DNA exon sequences were aligned with the MACSE software, which aligns coding DNA sequences using their amino-acid translation while allowing for potential frame shifts (Ranwez et al. 2011). After this multiple alignment step, we devised the following procedure to remove the so-called "outlier" sequences, *i.e.*, misaligned, paralogous or illegitimately-captured sequences. A maximum likelihood (ML) tree was inferred from each exon alignment using RAxML v8.2.8 (Stamatakis 2014) under the GTR+Γ model of DNA sequence evolution (Rodríguez et al. 1990; Yang 1996). In each tree, patristic distances between each taxon and the ancestral node of the crown ingroup (all Echimyidae and Capromyidae) were computed. As "outlier" sequences are likely to result into a topological misplacement and longer branches leading to it, we thus analyzed variations in the patristic distances with R (R. Core Team 2018). We used a linear model that included two independent qualitative predictor variables: the exon (with 483 modalities) and the species (41 modalities) involved in the corresponding log-transformed root-to-tip distance. We then flagged as "outlier" every sequence involved in a root-to-tip distance with a residual value of the model greater than 3 standard deviations. Outlier sequences were then pruned out from corresponding alignments. After this curating step, sequences were re-aligned with MACSE using default parameters, and sites with more than 20% of ambiguous character states and gaps were removed using trimAl v1.3. (Capella-Gutiérrez et al. 2009).

The nucleotide and amino-acid datasets can be accessed in the following GitHub repository: https://github.com/CourcelleM/Courcelle_et_al2018_Echimyidae

Zipped nexus files contain the raw and curated alignments of 483 exons for the 40 species studied.

2.5. Phylogenetic analyses

All curated exon alignments were concatenated leading to a supermatrix with 601,816 sites, comprising 14 % missing character states. At the protein level, the supermatrix comprised 194,088 amino acid sites, with 9% missing character states. Previous analyses highlighted short divergence times and fast molecular evolution of some Echimyidae clades (Lara et al. 1996; Fabre et al. 2013; 2017). For example, the uncertain position of the long branch of *Carterodon sulcidens* destabilized the species tree of Fabre et al. (2017). An evolutionary model able to efficiently deal with homoplasies and long-branch attraction was thus recommended in order to reconstruct the Echimyidae phylogeny. A Bayesian inference was conducted on the concatenated exon alignments, both at the nucleotide and amino-acid levels, using the CAT-GTR+Γ mixture model implemented in PhyloBayes-MPI v1.8 (Lartillot et al. 2009). This sequence evolution model takes across-site composition and substitution rates heterogeneity into account and is thus well suited to deal with long-branch attraction artifacts (Lartillot and Philippe 2004). Two Markov Chains Monte Carlo (MCMC) were run independently and parameters were sampled every 5 cycles. Chain convergence was assessed when the two chains inferred the same bipartitions frequencies. Node posterior probabilities were computed after respectively 1,210 and 1,245 cycles (corresponding to approximately 1,000,000 generations) with a 30% burn-in discarding the first 400 trees.

A potential limitation of the supermatrix approach is that the phylogenetic signal brought by each specific gene might be diluted during the joint analyses (Song et al. 2012). This can lead to incongruences when the genes bear conflicting information, for example due to incomplete lineage sorting or hidden paralogy (Kubatko and Degnan 2007; Kuraku 2013). We thus also estimated the echimyid species tree by a multi-coalescent approach. To account for phylogenetic uncertainty, 100 bootstrap trees were inferred for each exon under the GTR+Γ model with RaxML. The resulting collection of 48,300 bootstrapped gene trees were then inputed to the ASTRAL-2 software (Mirarab and Warnow 2015). As "gene+site -wise" bootstrapping may generate artefactual conflicts among gene trees (Simmons et al. 2019), we repeated this analysis using the "--gene-only" bootstrapping

option of ASTRAL-III (Zhang et al. 2018). An assumption of summary coalescence methods like ASTRAL is that there is no intra-locus recombination. According to Scornavacca and Galtier (2017), the exon-level at which we analysed our dataset is likely the most robust for phylogenomic studies. To ascertain that intra-locus recombination did not distort our results, we re-analyzed our dataset using the SVDquartets method (Chifman and Kubatko 2015) as implemented in PAUP* (Swofford 2002). As SVDquartets uses quartets of taxa to infer phylogenetic relationships at the individual alignment site-level, it does not require the assumption that there is no intralocus recombination.

A corollary point relative to our analysis protocol is the impact level on phylogenetic inferences of the large spectrum of evolutionary rates, α values and nucleotide compositions of exons sampled (see Betancur-R et al. 2013, Townsend and Naylor 2007, and Romiguier et al. 2013 respectively). We therefore measured the congruence between the reference supermatrix tree and phylogenetic trees inferred along a molecular evolution property gradient. Exons were sorted out according to their (1) rate of evolution, (2) α value and (3) nucleotide composition and were then concatenated. This ensured us to obtain three collections of aligned exons with ascending values of RER, α and GC3 respectively. We then used consecutive bins of 15,000 sites from each matrix to analyze datasets of the same length but with gradually contrasted molecular properties. We inferred 100 bootstrap trees for each bin and computed their distance to the "reference" tree using the K-score metric, which takes into account the topological and branch length differences (Soria-Carrasco et al. 2007). As we are interested in the K-score difference between the bins, the choice of the reference tree was not of prime importance, and we conventionally chose to use the phylogenetic tree yielded by our supermatrix Bayesian analysis.

2.6. Molecular dating

Molecular dating analyses were performed with PhyloBayes v4.1 (Lartillot et al. 2009). We used a Bayesian relaxed molecular clock with rate autocorrelation along the branches of the tree (Thorne et al. 1998; Thorne and Kishino 2002) on the reference topology yielded by the supermatrix analysis. Five uniform fossil constraints were employed to calibrate the node of the phylogeny, with soft bounds to allow dates to be drawn in an interval 5 % greater than the set boundaries (Yang and Rannala 2006). The younger bound of the root was set to the oldest hystricognaths of South America at 41.0 Million years ago (Mya) (Antoine et al. 2012). The older

bound was set to the early Eocene at 56.0 Mya. In the outgroup, the divergence time among Chinchilloidea and Octodontoidea was set between the oldest stem Chinchilloidea member *†Eoviscaccia frassinettii* (31.3 Mya, Bertrand et al. 2012) and the earliest caviomorph fossils (41.0 Mya, Antoine et al. 2012). In the ingroup, the first calibration used was the divergence between *Trinomys* and *Clyomys+Euryzygomatomys* genera (Euryzygomatomyinae subfamily). Based on the *†Theridomysops parvulus* age, the oldest stem lineage to *Euryzygomatomys+Clyomys* genera, a lower bound was set to 6.0 Mya for the Euryzygomatomyinae node (Verzi et al. 2013). The upper bound was subsequently set to 11.8 Mya, the deepest level of the Laventan SALMA (South American Land Mammal Age) (Verzi 2008). We then constrained the most recent common ancestor of the Myocastorini to range between the oldest stem taxon to *Thrichomys* (6.0 Mya, Olivares et al. 2012; Verzi et al. 2013) and the start of the Laventan SALMA (11.8 Mya, Verzi et al. 2008). Finally, we dated the lower bound of the Echimyini divergence with the oldest representative of the *Echimys+Phyllomys* clade : *†Maruchito trilofodonte* (15.7 Mya, Vucetich et al. 1993; Kay et al. 1997). The upper bound was set to the start of the Deseadan SALMA (24.2 Mya, Dunn et al. 2013).

2.7. Locomotor habits reconstruction

The evolution of locomotor habits of echimyids was modeled using the R package *diversitree* (FitzJohn 2012). Coding for locomotion lifestyle was as follows: (1) semi-arboreal, scansorial, or arboreal (*Callistomys, Capromys, Dactylomys, Echimys, Isothrix, Kannabateomys, Mesocapromys, Makalata, Pattonomys, Plagiodontia, Mesomys, Mysateles, Olallamys, Phyllomys)*; (2) semi-fossorial (*Clyomys, Euryzygomatomys, Carterodon*); (3) terrestrial (*Proechimys, Geocapromys, Hoplomys, Thrichomys, Trinomys*); and (4) semi-aquatic (*Myocastor*). Ancestral character state estimations were implemented in the diversitree package in the R programming language (R. Core Team 2018) using the “make.mkn” function. This maximum likelihood (ML) mapping approach fits the model of Pagel (1994) to reconstruct states at all ancestral nodes of our phylogeny using the Mk2/Mk-n models. Three models of trait evolution were fitted and compared: (i) all rates for locomotor habit transitions are equal, (ii) rates are symmetrical, and (iii) all rates are allowed to be different. The first model provided the lowest AIC. Plotting of the results were performed using a R function from Moen et al. (2015).

3. RESULTS AND DISCUSSION

3.1. Exon capture as an efficient tool for phylogenomic analysis

In this study, we performed an exon capture experiment in order to collect a vast molecular dataset for species for which previous analyses were limited to mitochondrial genomes and a handful of nuclear exons (Lara et al. 1996; Huchon and Douzery 2001; Leite and Patton 2002; Upham and Patterson 2015; Fabre et al. 2017; Álvarez et al. 2017). Given a reference sequence, exon capture allows to easily isolate and sequence DNA from other individuals from the same species, but its prime interest is to extend this capacity to genes from sister-group taxa that are likely to exhibit more sequence divergence. If a close enough reference is available for non-model species, this technique is a viable alternative to shotgun sequencing for many investigations where molecular data is required, from ancient DNA studies to whole-genome sequencing and phylogenomics (Enk et al. 2014; McCormack et al. 2016; Teasdale et al. 2016). The capture approach is limited by the raw genetic distance between reference and target sequences: hybridization efficiency decreases when probes increasingly differ from target sequences, resulting into reduced efficiency of captures and amplifications. The protocol followed in the present study allowed us to capture exon sequences from Echimyidae and Capromyidae using an octodontid representative (here *Octodon degus*), for sequence divergence up to 20% (raw DNA pairwise distance from *O. degus* to other species, computed from final exon alignments). This is in agreement with other studies that reported capture enrichment success for divergent sequences at least up to 15% (Bragg et al. 2016), and up to 39% with modified hybridization protocols (Li et al. 2013). We thus validate this exon capture protocol for two families whose divergence time is estimated around 23.6 Mya (Upham and Patterson 2015), 31 Mya (Álvarez et al. 2017), or 21.1 Mya ([25.1 – 18.1] Mya in the present analysis). The number of paired raw reads captured for each of the 35 echimyid taxa sequenced here ranged from 119,544 to 10.4 Million (Table 1). Inter-species read number heterogeneity is often explained by a lower quality of the biological material because degraded DNA is more difficult to sequence and assemble through new generation sequencing (Burrell et al. 2015). This is not the case here, as the number of raw reads was not correlated to the age of tissues (Spearman's rank correlation coefficient, p-value = 0.4), nor explained by the origin of the sample tissue (ANalysis Of VAriance, p-value = 0.16), two factors known to hamper NGS results (Almeida et al. 2014). Despite variation in the number of reads available for each taxon, 89 to 99% of the 483 targeted exons were captured for each species (Table 1). Moreover, our methodology allowed us to capture a sample of nuclear exons covering a wide

range of molecular and evolutionary characteristics (Fig. 1). The relative evolutionary rates (RER) of exons computed from the gene trees inferences ranged between approximately 0.01 (very conserved) and 6 (fast-evolving), denoting a 600-fold difference in the exons rates of evolution. The α parameter varies from 0.02 to 13.01. The highest value corresponds to the first exon of the gene TEX10 (Ensembl reference ENSG00000136891). It was a clear outlier as the second highest α was estimated at approximately 2.2. Checking the alignment of TEX10 revealed that the sequence is extremely conserved (RER of 0.01), but rare substitutions seemed to occur at random third codon positions along the sequence, leading to an almost uniform probability of substitution for each site (very high α value). It is interesting to note that the distribution of α and GC3 values show respectively 82% and 86% overlap between mammalian and rodent scales (Fig. 1B and 1C). It thus seems that the estimates of α and GC3 values do not significantly depend upon the taxonomic scale at which they are computed. Values from the OrthoMam database therefore seem to be a reasonable approximation for mammalian species at a lower taxonomic scale. Relative evolution rates however are found generally lower in our echimyids dataset than in OrthoMam (63% distribution overlap, Fig. 1A). Since our dataset contains phylogenetically closer taxa than those in the Orthomam database, exons show less variability among the species studied here, which leads to lower relative evolutionary rates.

Even if our results highlight a broad range of exon capture success, we did observe a negative correlation between assembly coverage and the relative rate of evolution of captured exons (Fig. 2B). This may be a hint that divergent sequences are harder to capture: the more they differ from the O. degus reference, the less efficient is the hybridization between probes and target sequences (Fig. 2A and 2B). Moreover, optimal exon coverage was expected for a balanced nucleotide content, *i.e.* a GC3 content of 50% (see *e.g.* Bi et al. 2012). Indeed, a strong nucleotide bias lowers sequence complexity, and therefore the probes specificity. Additionally, amplification steps are less efficient when sequences are richer in G and C nucleotides (Viswanathan et al. 1999). In this study, mean exon coverage strictly decreased with GC3 content (Fig. 2C). This relation suggests that the lower complexity of probes in AT-rich regions may be less detrimental than the difficulties brought by G+C-rich sequences, such as stronger DNA secondary structure hampering amplification steps (Kozarewa et al. 2009), or higher probability of sequencing error (Nakamura et al. 2011). Thus, the A+T-richness of sequences does not seem to have been a limiting factor to capture exon at the rodent family taxonomic scale, as illustrated by the higher sequencing depth of A+T-rich exons (Fig. 2C).

3.2. A fully resolved Echimyidae phylogeny

The phylogeny of South American spiny rats greatly benefited from the advent of high-throughput sequencing and molecular phylogeny. Morphological analyses (Carvalho and Salles 2004; Emmons 2005) and molecular star-phylogenies (Lara et al. 1996; Leite and Patton 2002) were followed by probabilistic methods using more sophisticated evolutionary models which, even if applied to only a handful of mitochondrial and nuclear genes, allowed a partial resolution of the Echimyidae phylogeny (Galewski et al. 2005; Upham and Patterson 2015; Fabre et al. 2013 ; 2017; Álvarez et al. 2017; Emmons and Fabre 2018). Still, several nodes were poorly supported or did not gain maximal statistical support (Fabre et al. 2017), despite the use of complete mitogenomes and up to five nuclear exons. Here, two independent phylogenetic approaches were used on our large nuclear dataset to untangle the Echimyidae higher-level phylogeny: (1) a Bayesian inference on the supermatrix of all exon sequences and (2) multicoalescent-based approaches using the gene trees as a primary input. The two analyses yielded highly similar results (Fig. 3): the same topology, and each node unambiguously supported, with respectively posterior probabilities and local posterior probabilities values equals to 1 (Fig. 3). The tree obtained is the same regardless of the bootstrapping approach used ("gene+site -wise" or "gene-wise", see section 2.5). The SVDquartets analysis also produced the same topology. Thus, we here report the first fully and well-supported genus-level Echimyidae phylogeny.

The paraphyly of Echimyidae as currently understood is evidenced as it includes the Capromyidae as sister group to the Brazilian echimyid *Carterodon.* We also recovered three major clades : (1) Capromyinae (new rank), comprising living hutias from the West Indies (including *Plagiodontia, Geocapromys, Capromys, Mysateles* and *Mesocapromys*), (2) Euryzygomatomyinae including terrestrial and semifossorial taxa from eastern and central South America (*Trinomys, Clyomys* and *Euryzygomatomys*), and (3) Echimyinae, comprising the remaining, mostly arboreal spiny rats (*Dactylomys, Echimys, Isothrix, Kannabateomys, Makalata, Mesomys, Ollalamys, Pattonomys, Phyllomys, Callistomys, Hoplomys, Myocastor, Proechimys* and *Thrichomys*). These findings reinforce phylogenetic hypotheses proposed in previous studies (Galewski et al. 2005; Upham and Patterson 2015; Fabre et al. 2017), however the relative positions of the three subfamilies were conflictual. In line with most recent phylogenies (Fabre et al. 2014; 2017), our results firmly corroborate sister-relationships between Capromyinae and *Carterodon* (for which we

create a new subfamily, see below), forming a clade sister to Euryzygomatomyinae. This clade is in turn sister to Echimyinae (Fig. 3).

Within Echimyinae, we again retrieved the surprising affinity between the semi-aquatic *Myocastor coypus* and the arboreal *Callistomys pictus* (Loss et al. 2014; Fabre et al. 2017). These taxa form a well-supported clade together with *Thrichomys* and *Hoplomys* + *Proechimys*, named Myocastorini *sensu* Fabre et al. (2017). Other arboreal genera of the subfamily are grouped in the Echimyini tribe. The lineage leading to the brush-tailed rat *Isothrix* diverged first, followed by the lineage leading to *Mesomys*. This finding is at odds with most previous studies that proposed a clade *Isothrix* + *Mesomys* (Fabre et al. 2017; Álvarez et al. 2017). We recovered a rather short branch between these two genera (approximately 0.002 substitutions per site) that could not be reliably inferred using only the small gene sampling of previous studies (Upham and Patterson, 2015; Fabre et al. 2017). The bamboo rat clade is also retrieved with *Kannabateomys amblyonyx* being a sister-taxon to a clade formed by *Dactylomys dactylinus* and *Olallamys albicauda* as in previous studies of Galewski et al. (2005) and Upham et al. (2012). Species of *Echimys* are closely related to *Phyllomys*, and this clade is sister to *Makalata,* forming another clade sister to *Pattonomys* in accordance with recent phylogenetic results (Fabre et al. 2017; Fabre and Emmons 2018). The subfamily Euryzygomatomyinae comprises *Trinomys* and its semifossorial sister-clade *Clyomys laticeps* and *Euryzygomatomys spinosus*. The Capromyinae includes *Capromys*, *Mesocapromys*, and *Mysateles*, which form the Capromyini tribe. This clade is the sister group to the Hispaniolan hutia *Plagiodontia aedium* (Plagiodontini).

As our phylogenetic analyses were performed on a wide sampling of nuclear exons with contrasting molecular properties, we evaluated whether the evolutionary rate, the substitution distribution and the G+C-content of exons distorted gene tree inferences. The variation of the congruence of trees reconstructed along a gradient of these three properties has been measured by the K-score metric (Fig. 4). We did not detect any clear impact of the molecular evolutionary rate, the value of the α parameter or the GC content on the level of the individual exon tree congruence. We however observed an edge effect where trees inferred from the subsamples with (1) low Relative Evolution Rates (RER) and (2) low α values show higher K-score values than the average (Fig. 4). This suggests that the slowest-evolving exons, as well as those where multiple substitutions occur, are more likely to generate homoplasy. A closer examination of the phylogenetic trees revealed that rather than providing strongly incongruent relationships among Echimyidae, these

samples lack resolution to resolve the shortest branches, such as the deep branching of the different subfamilies or the branches leading to *Carterodon sulcidens* or *Mesomys* / *Lonchothrix* (Fig. 3). As these issues are compensated by other exons in our dataset, we postulate that including contrasting molecular characteristics did not mislead our phylogenetic inferences but instead allowed us to reach such a high statistical resolution.

3.3. Taxonomic implications for the Capromyidae and *Carterodon sulcidens*

With the resolution of the phylogenetic relationships among the Echimyidae clades, this study brings the first unequivocal material to disentangle the higher-level echimyid classification. First, the relationships of the Caribbean hutias, the Capromyidae, in regard to Echimyidae have been uncertain for a long time (Wood and Howland 1979). Molecular studies yielded two alternative topologies that question the taxonomic status of Capromyidae as well as the monophyly of Echimyidae (see Fabre et al. 2017). The first hypothesis was the reciprocal monophyly of Capromyidae and Echimyidae (Upham and Patterson 2012). This view strengthened the status of Capromyidae as a distinct family, closest relative of extant echimyids. The second hypothesis recovered in the literature consisted in grouping Capromyidae and Euryzygomatomyinae within Echimyidae (Galewski et al. 2005; Fabre et al. 2013; 2014; 2016). Echimyidae would here become paraphyletic, unless Capromyidae are actually viewed as a subfamily within Echimyidae.

Our results provide unambiguous support to the second hypothesis, i.e. a Capromyinae + Euryzygomatomyinae clade nested within Echimyidae (Fig. 3, Fig. 5). In the present study, we used both a wide taxonomic sampling, including all extant genera of Capromyidae, as well as the best character sampling of Echimyidae and Capromyidae rodents to date with more than 600,000 nuclear DNA and 200,000 protein sites. We also used slow-evolving exons with various molecular evolution characteristics (see previous parts) and different phylogenetic methodologies to decrease the impact of well-known methodological artifacts such as site saturation, long-branch artifacts (Lartillot and Philippe 2004; Delsuc et al. 2005), as well as incomplete lineage sorting and hidden paralogy (Kubatko and Degnan 2007; Kuraku 2013). To fit our results as well as several recent phylogenetic inferences (Fabre et al. 2014; 2017) in a systematic framework, we thus propose to rank Capromyidae, as defined by Smith (1842), as a subfamily – Capromyinae – within Echimyidae definition *sensu* McKenna and Bell (1997). Moreover, previous studies struggled to assert the

phylogenetic placement of the enigmatic *Carterodon sulcidens* (but see also Vucetich and Verzi 1991), also referred to as *incertae sedis* (Fabre et al. 2016; 2017). We actually provide here the first highly supported relationship of this semifossorial Brazilian species within the echimyid phylogeny. Both the supermatrix and multi-coalescent analyses place *C. sulcidens* as a deep-branching sister-group to the Capromyinae. The clade formed by the Carterodontinae and the Capromyinae is defined by 175 exclusive amino-acid synapomorphies. Examples of exclusive synapomorphies with no missing taxa in our dataset include i) a Valine to Glutamic acid substitution at position homologous to site 887 of the human Protocadherin Fat 1 protein (UniprotKB: Q14517); ii) a Leucine to Serine substitution at position homologous to site 3009 of the human Apolipoprotein B protein (UniprotKB: P04114).

This finding is inconsistent with previous studies where *C. sulcidens* was more closely related to Euryzygomatomyinae (Emmons 2018; Upham and Patterson 2015; Álvarez et al. 2017). Here *C. sulcidens* displays the longest isolated branch from the phylogeny, and its path to the Echimyidae ancestral node is 20 % longer than the mean for other species (Fig. 3). Its new relationship implies major biogeographical consequences for the Greater Antilleans hutias as well as convergent evolution of fossoriality in echimyids (see also Carvalho and Salles 2004; Candela and Rasia 2012; Fabre et al. 2013). Indeed, it has been shown that ecological shifts were common in Echimyidae, leading to unexpected relationships such as the clade formed by the arboreal *Callistomys* and the semi-aquatic *Myocastor* in a mostly terrestrial tribe (Loss et al. 2014; Fabre et al. 2017). The phylogenetic position of *Carterodon sulcidens* implies that the semifossorial adaptations likely evolved at least twice in Echimyidae during the Middle-Miocene (Table 2, Figure S2): once along the branch leading to the *Euryzygomatomys* + *Clyomys* clade, and once along the branch of *Carterodon*. This lifestyle has been suggested to be an adaptation in terms of thermoregulation as well as a way to escape predators in the open environments of Cerrado and Caatinga (de Albuquerque et al. 2012), the South-American biomes which host all semifossorial Echimyidae species and several other semifossorial caviomorph rodents (Fabre et al. 2017). This lifestyle convergence with its related morphological adaptations may have misled early taxonomists in grouping together *Carterodon sulcidens* and semifossorial Euryzygomatomyinae on the basis of morphological characters (Carvalho and Salles 2004; Emmons 2005; Candela and Rasia 2012). However, some morphologists have already pointed the distinctiveness of *Carterodon* with its unique grooved upper incisors (Fig. 3, insert A) (Bishop 1974), and some autapomorphic cranio-

mandibular characters (Vucetich and Verzi 1991). Taking together these molecular, biogeographical and morphological evidences, we propose to define a new subfamily Carterodontinae, which would comprise the extant monotypic genus *Carterodon* (Fig. 3):

Family Echimyidae, Gray, 1825

- Carterodontinae new subfamily

Type genus: *Carterodon* Waterhouse, 1848. Included genera: only the monotypic *Carterodon*, represented by *Carterodon sulcidens* (Lund, 1841). Diagnosis: small (<200g adult body weight, Bezerra et al. 2011), semifossorial rodent with short tail (ca. 50% head and body length); anterior face of the upper incisor grooved, protoloph absent in upper molariforms (Patton and Emmons 2015); canal of the infraorbital nerve in the infraorbital foramen absent (Carvalho and Salles 2004). Distribution: central and western part of Brazil in the Cerrado biome, ranging from the states of Minas Gerais to Mato Grosso (Bezerra et al. 2011).

Based on the results of this study, we propose a revised classification of living echimyid rodents above the species level.

Mammalia Linnaeus, 1758
- Rodentia Bowdich, 1821
  - Octodontoidea Waterhouse, 1839
    - Echimyidae Gray, 1825
      - Capromyinae Smith, 1842
        - Capromyini Smith, 1842
          - *Capromys* Desmarest, 1822
          - *Geocapromys* Chapman, 1901
          - *Mesocapromys* Varona, 1970
          - *Mysateles* Lesson, 1842
        - Plagiodontini Ellerman, 1940
          - *Plagiodontia* F. Cuvier, 1836
      - Carterodontinae Courcelle, Tilak, Leite, Douzery & Fabre, **subfam. nov.**
        - *Carterodon* Waterhouse, 1848
      - Euryzygomatomyinae Emmons, 2005
        - *Clyomys* Thomas, 1916
        - *Euryzygomatomys* Goeldi, 1901
        - *Trinomys* Thomas, 1921
      - Echimyinae Gray, 1825
        - Myocastorini Ameghino, 1902
          - *Callistomys* Emmons & Vucetich, 1998
          - *Hoplomys* J. A. Allen, 1908
          - *Myocastor* Kerr, 1792
          - *Proechimys* J. A. Allen, 1899
          - *Thrichomys* Trouessart, 1881
        - Echimyini Gray, 1825
          - *Dactylomys* I. Geoffroy St.-Hilaire, 1838
          - *Diplomys* Thomas, 1916
          - *Echimys* Cuvier, 1809
          - *Isothrix* Wagner, 1845
          - *Kannabateomys* Jentink, 1891
          - *Leiuromys* Emmons & Fabre, 2018
          - *Lonchothrix* Thomas, 1920
          - *Makalata* Husson, 1978
          - *Mesomys* Wagner, 1845
          - *Olallamys* Emmons, 1988
          - *Pattonomys* Emmons, 2005
          - *Phyllomys* Lund, 1839
          - *Santamartamys* Emmons, 2005

Toromys Iack-Ximenes, de Vivo & Percequillo, 2005

3.4.Biogeographical and systematic perspectives

The Echimyidae phylogeny inferred from ca. 500 exons also bolsters the current hypothesis of the echimyid dispersal, with the Caribbean Capromyinae diverging from Neotropical lineages during the mid-Miocene (Fabre et al. 2014; 2016). Mid-Miocene is marked by the opening of the habitats in South America which have shaped Echimyidae diversification along with the emergence of open habitat specialists such as *Carterodon*, *Clyomys*, and *Thrichomys* (Galewski et al. 2005; Fabre et al. 2013; 2016). This period of echimyid ecological and morphological diversifications is also accompanied by the diversification of the hutias, thanks to their arrival in the Greater Antilles. This colonization seems to have occurred in the mid-Miocene as inferred by our molecular dating analysis (Fig. 5, Table 2) as well as prior dating of this subfamily (Fabre et al. 2013; 2014; Upham and Patterson 2015). The close affinity of hutias with *Carterodon* might imply a potential origin in central South America. Unnamed capromyids have been reported from Quebrada Honda, a middle Miocene site in Bolivia (McKenna and Bell 1997), but none of the rodents from this site has been described in detail, and recent changes in taxonomy indicate that capromyids do not appear to be present there (Croft 2007). Extinction has probably hindered the original continental source pool of hutia ancestors, letting *Carterodon* as the sole echimyid survivor of their ancient radiation. An interesting avenue is that several echimyid subfossils (e.g. Heteropsomyinae) are known from the Greater Antilles deposits (Van der Geer et al. 2011). Some of these lineages (*Boromys*, *Brotomys*) share with *Carterodon* a simple occlusal molar pattern (Nowak and Walker 1999). Further investigations of morphological and molecular systematic of *Carterodon* as well as Caribbean rodent subfossils will be paramount. Such analysis will help us estimate the number of colonizations to the West Indies, as well as shed light on the extinct morphological disparity of the diverse Echimyidae.

Acknowledgments

We thank Fabien Condamine, François Catzeflis, Frédéric Delsuc, Nicolas Galtier, Vincent Ranwez, Nathan Upham and two anonymous referees for discussions about analyses and the redaction of this article. We are grateful to the following people and institutions for granting access to study skins and samples: Darrin Lunde, Nicole Edmison, and Louise Emmons (NMH, Smithsonian Institution), Chris Conroy and Jim Patton (MVZ Berkeley), Ana Carolina Loss and

Leonora Costa (UFES), Judy Chupasko and Hopi Hoekstra at the Museum of Comparative Zoology (MCZ, Cambridge USA), Steve van Der Mije (Naturalis Biodiv Center, Leiden, Netherland), Burton Lim (Royal Ontario Museum Toronto, Canada), and François Catzeflis (UM-ISEM). We also thank Alison Devault (Arbor Biosciences) for her help on the capture kit design. We acknowledge the financial support of the "Agence Nationale de la Recherche" through the GAARanti project grant (ANR-17-CE31-0009). P.-H.F. acknowledges the SYNTHESYS Foundation for funding his work in the BMNH collections (GB-TAF-2735, GB-TAF-5026, and GB-TAF-5737). P.-H.F. also obtained financial support from Agence Nationale de la Recherche (CEBA, ref. ANR-10-LABX- 25-01). YLRL had continuous support from Fundação de Amparo a Pesquisa e Inovação do Espírito Santo (FAPES) and Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), Brazil. The bio-informatics analyses of this article largely benefited from the Montpellier Bioinformatics Biodiversity cluster computing platform. This publication is the contribution n°XXX-XXX SUD of the Institut des Sciences de l'Évolution de Montpellier (UMR 5554 UM-CNRS-IRD-EPHE).

Authors declare no competing interests.

TABLES

Table 1 Capture and sequencing statistics of the 483 nuclear exons for the 35 studied Echimyidae species

Species	Raw reads	Unique reads	Unique reads percentage	Captured exons
Callistomys pictus	10 446 228	1 258 230	12.04	477
Capromys pilorides	1 302 422	269 201	20.67	461
Carterodon sulcidens	6 165 400	1 466 951	23.79	476
Clyomys laticeps	6 828 172	1 657 105	24.27	473
Dactylomys dactylinus	3 233 010	1 118 435	34.59	462
Echimys chrysurus	119 544	33 143	27.72	430
Echimys saturnus	5 545 284	984 119	17.75	468
Euryzygomatomys spinosus	7 079 304	1 723 493	24.35	433
Geocapromys thoracatus	7 927 390	2 013 689	25.40	477
Hoplomys gymnurus	8 108 526	1 926 392	23.76	472
Isothrix pagurus	5 915 678	1 958 259	33.27	474
Isothrix sinnamariensis	9 330 248	1 595 364	17.10	473
Kannabateomys amblyonyx	6 632 906	1 293 378	19.50	472
Makalata didelphoides	3 587 512	910 757	25.39	472
Mesocapromys melanurus	3 220 430	768 595	23.87	478
Mesocapromys nana	7 632 848	1 972 264	25.84	476
Mesomys hispidus	9 871 930	1 983 491	20.09	478
Myocastor coypus	364 238	142 555	39.14	459
Mysateles prehensilis	1 536 955	646 248	42.05	475
Olallamys albicauda	10 053 242	1 923 423	19.13	468
Pattonomys carrikeri	7 661 732	1 732 204	22.61	475
Phyllomys blainvillii	3 233 902	583 265	18.04	467
Phyllomys dasythrix	5 140 752	991 961	19.30	475
Phyllomys lundi	8 579 706	1 941 857	22.63	479
Phyllomys mantiqueirensis	4 427 564	802 854	18.13	478
Phyllomys pattoni	3 201 342	654 229	20.44	476
Plagiodontia aedium	690 040	555 914	80.56	474
Proechimys cuvieri	7 405 714	1 875 484	25.32	474
Thrichomys apereoides	9 789 590	1 793 456	18.32	437
Trinomys albispinus	168 264	33 830	20.11	438

Species	Raw reads	Unique reads	Unique reads percentage	Captured exons
Trinomys iheringi	9 358 598	1 645 783	17.59	470
Trinomys panema	5 893 146	4 774 192	81.01	478
Trinomys paratus	7 329 946	1 355 485	18.49	474
Trinomys setosus	2 113 354	429 744	20.33	472
Trinomys yonenagae	4 869 508	1 087 080	22.32	477

Table 2 Molecular dating of deep Echimyidae phylogenetic nodes in this analysis as compared to recent studies. Node labels are reported on Fig. 5. Node ages and 95% credibility intervals are given in million years ago.

Node	This study	Fabre et al. 2017	Álvarez et al. 2017	Upham and Patterson 2015
A	14.4 [18.8-13.2]	13.6	26.73	18.2
B	14.1 [18.2-12.9]	12.7	25.36	17.4
C	13.4 [16.1-12.3]	NA	NA	NA
D	8.6 [10.1-7.5]	7.6	15.56	9.8
E	12.8 [14.0-11.7]	12.7	23.65	15.8
F	13.1 [17.1-12.0]	13.3	25.05	17.1
G	11.8 [15.7-10.7]	12.0	23.26	16.2
H	10.8 [11.9-9.8]	9.4	22.49	15.5

FIGURES CAPTION

Fig. 1 Variation of molecular properties and evolution characteristics of the 483 target nuclear exons, as captured in this study (histogram) at the hystricognath rodents scale and according to the Orthomam reference at the mammals scale (dashed line distribution). (A) Relative evolutionary rates of exons, as computed from the RaxML trees using a Super Distance Matrix approach. Values range from 0.01 to 6.4. (B) Value of the shape parameter α of the Gamma distribution for the among-site substitution rate heterogeneity, estimated by maximum-likelihood during the inference of RaxML trees for each exon alignment. Values range from 0.02 to 2.2. One outlier data point corresponding to the first exon of the gene TEX10 (Ensembl reference ENSG00000136891) is not reported (see Results and Discussion). (C) Mean G+C content in the 3rd codon position of exon's alignments. Values range from 0.24 to 0.96.

Fig. 2 Variation in sequencing coverage (reads per site) of the 483 captured nuclear exons. Median coverage was computed for each exon from the sequencing depth of the 36 Echimyidae species. (A) Coverage of exons as a function of the mean distance from *O. degus* to Echimyidae species in the corresponding alignment. (B) Coverage of exons according to the exon's relative rate of evolution. (C) Coverage of exons according to the mean G+C content in the 3rd codon position of the corresponding alignment.

Fig. 3 Phylogenetic tree of Echimyidae species inferred from a Bayesian analysis of the concatenation of 483 nuclear exons (601,816 DNA sites). Sequence evolution was described under the CAT + GTR + Γ mixture model. Posterior probabilities are reported for nodes of the tree. The phylogeny is rooted with the Octodontidae *Octodon degus*. For readability, other Hystricognathi outgroups have been pruned. Branch lengths are expressed in expected number of substitutions per site. Colors delimiting clades are given along the topology. Insert **(a)** represents a ventral view of the skull of *Carterodon sulcidens*. The arrow highlights the distinctive grooved incisors of this species.

Fig. 4 K-score metric between the reference phylogenetic tree of Echimyidae (Fig. 3) and 40 trees inferred from 15,000 sites-long subsamples of our dataset with a gradient of Relative Rate of Evolution (A), a gradient of among-site rate variation α (B) and a gradient of G+C content on third codon positions (C). X-axes for these figures are not linear scales. Bins boxes are arbitrarily evenly spaced. Parameter values for each five bins are annotated. Values range approximately from 0.36 to 2.6 for the RER, 0.064 to 0.992 for α and 0.34 to 0.67 for the GC3.

Fig. 5 Molecular dating of the Echimyidae family divergence. The topology of the chronogram was constrained to the result of our Bayesian analysis of the concatenation of 483 nuclear exons (Fig. 3). Green-colored bars represent the 95% credibility time interval inferred for nodes. The gray overlay displays the expected time window of the West Indies colonization by the Capromyinae subfamily. It has been set between the upper bound of node C (origin of Carterodontinae) and lower bound of the node D (diversification of Capromyinae). The timescale is expressed in million years.

REFERENCES

de Albuquerque UP, de Lima Araújo E, El-Deir ACA, de Lima ALA, Souto A, Bezerra BM, Ferraz EMN, Maria Xavier Freire E, Sampaio EV de SB, Las-Casas FMG. 2012. Caatinga revisited: ecology and conservation of an important seasonal dry forest. Sci. World J. 2012.

Almeida FC, Giannini NP, Simmons NB, Helgen KM. 2014. Each flying fox on its own branch: A phylogenetic tree for *Pteropus* and related genera (Chiroptera: Pteropodidae). Mol. Phylogenet. Evol. 77:83–95.

Álvarez A, Moyers Arévalo RL, Verzi DH. 2017. Diversification patterns and size evolution in caviomorph rodents. Biol. J. Linn. Soc. 121:1–16.

Antoine P-O, Marivaux L, Croft DA, Billet G, Ganerød M, Jaramillo C, Martin T, Orliac MJ, Tejada J, Altamirano AJ, et al. 2012. Middle Eocene rodents from Peruvian Amazonia reveal the pattern and timing of caviomorph origins and biogeography. Proc. R. Soc. B 279:1319–1326.

Bejerano G, Pheasant M, Makunin I, Stephen S, Kent WJ, Mattick JS, Haussler D. 2004. Ultraconserved Elements in the Human Genome. Science 304:1321–1325.

Bertrand OC, Flynn JJ, Croft DA, Wyss AR. 2012. Two New Taxa (Caviomorpha, Rodentia) from the Early Oligocene Tinguiririca Fauna (Chile). Am. Mus. Novit.:1–36.

Betancur-R R, Li C, Munroe TA, Ballesteros JA, Ortí G. 2013. Addressing gene tree discordance and non-stationarity to resolve a multi-locus phylogeny of the flatfishes (Teleostei: Pleuronectiformes). Syst. Biol. 62:763–785.

Bezerra AMR, Marinho-Filho J, Carmignotto AP. 2011. A review of the distribution, morphometrics, and habit of Owl's Spiny Rat *Carterodon sulcidens* (Lund, 1841)(Rodentia: Echimyidae). Zool. Stud. 50:566–576.

Bi K, Vanderpool D, Singhal S, Linderoth T, Moritz C, Good JM. 2012. Transcriptome-based exon capture enables highly cost-effective comparative genomic data collection at moderate evolutionary scales. BMC Genomics 13:403.

Bishop IR. 1974. An annotated list of Caviomorph rodents collected in North-Eastern Mato Grosso, Brazil. Mammalia 38:489–502.

Bohmann K, Evans A, Gilbert MTP, Carvalho GR, Creer S, Knapp M, Yu DW, de Bruyn M. 2014. Environmental DNA for wildlife biology and biodiversity monitoring. Trends Ecol. Evol. 29:358–367.

Bragg JG, Potter S, Afonso Silva AC, Hoskin CJ, Bai BYH, Moritz C. 2018. Phylogenomics of a rapid radiation: the Australian rainbow skinks. BMC Evol. Biol. 18:15.

Bragg JG, Potter S, Bi K, Moritz C. 2016. Exon capture phylogenomics: efficacy across scales of divergence. Mol. Ecol. Resour. 16:1059–1068.

Branstetter MG, Longino JT, Ward PS, Faircloth BC. 2017. Enriching the ant tree of life: enhanced UCE bait set for genome-scale phylogenetics of ants and other Hymenoptera. Methods Ecol. Evol. 8:768–776.

Brown WM, George M, Wilson AC. 1979. Rapid evolution of animal mitochondrial DNA. Proc. Natl. Acad. Sci. 76:1967–1971.

Burrell AS, Disotell TR, Bergey CM. 2015. The use of museum specimens with high-throughput DNA sequencers. J. Hum. Evol. 79:35–44.

Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL. 2009. BLAST+: architecture and applications. BMC Bioinformatics 10:421.

Candela AM, Rasia LL. 2012. Tooth morphology of Echimyidae (Rodentia, Caviomorpha): homology assessments, fossils, and evolution. Zool. J. Linn. Soc. 164:451–480.

Capella-Gutiérrez S, Silla-Martínez JM, Gabaldón T. 2009. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. Bioinformatics 25:1972–1973.

Carpenter ML, Buenrostro JD, Valdiosera C, Schroeder H, Allentoft ME, Sikora M, Rasmussen M, Gravel S, Guillén S, Nekhrizov G, et al. 2013. Pulling out the 1%: Whole-Genome Capture for the Targeted Enrichment of Ancient DNA Sequencing Libraries. Am. J. Hum. Genet. 93:852–864.

Carvalho GAS, Salles LO. 2004. Relationships among extant and fossil echimyids (Rodentia: Hystricognathi). Zool. J. Linn. Soc. 142:445–477.

Chifman J, Kubatko, L. 2014. Quartet inference from SNP data under the coalescent model. Bioinformatics 30:3317-3324.

Christmas MJ, Biffin E, Breed MF, Lowe AJ. 2016. Finding needles in a genomic haystack: targeted capture identifies clear signatures of selection in a nonmodel plant species. Mol. Ecol. 25:4216–4233.

Clay OK, Bernardi G. 2011. GC3 of Genes Can Be Used as a Proxy for Isochore Base Composition: A Reply to Elhaik et al. Mol. Biol. Evol. 28:21–23.

Croft DA. 2007. The middle Miocene (Laventan) Quebrada Honda fauna, southern Bolivia and a description of its notoungulates. Palaeontology 50:277–303.

Delsuc F, Brinkmann H, Philippe H. 2005. Phylogenomics and the reconstruction of the tree of life. Nat. Rev. Genet. 6:361–375.

Delsuc F, Kuch M, Gibb GC, Hughes J, Szpak P, Southon J, Enk J, Duggan AT, Poinar HN. 2018. Resolving the phylogenetic position of Darwin's extinct ground sloth (Mylodon darwinii) using mitogenomic and nuclear exon data. Proc R Soc B 285:20180214.

Delsuc F, Scally M, Madsen O, Stanhope MJ, Jong D, W W, Catzeflis FM, Springer MS, Douzery EJP. 2002. Molecular Phylogeny of Living Xenarthrans and the Impact of Character and Taxon Sampling on the Placental Tree Rooting. Mol. Biol. Evol. 19:1656–1671.

Douzery EJP, Scornavacca C, Romiguier J, Belkhir K, Galtier N, Delsuc F, Ranwez V. 2014. OrthoMaM v8: A Database of Orthologous Exons and Coding Sequences for Comparative Genomics in Mammals. Mol. Biol. Evol. 31:1923–1928.

Dunn RE, Madden RH, Kohn MJ, Schmitz MD, Strömberg CAE, Carlini AA, Ré GH, Crowley J. 2013. A new chronology for middle Eocene–early Miocene South American Land Mammal Ages. GSA Bull. 125:539–555.

Emmons LH. 2005. A Revision of the Genera of Arboreal Echimyidae (Rodentia: Echimyidae, Echimyinae), With Descriptions of Two New Genera. In: Mammalian diversification: from chromosomes to phylogeography. Vol. 133. University of California Publications in Zoology Series. Lacey E, Myers P. p. 247–310.

Emmons LH, Fabre P-H. 2018. A Review of the *Pattonomys*/*Toromys* Clade (Rodentia: Echimyidae), with Descriptions of a New *Toromys* Species and a New Genus. Am. Mus. Novit. 3894:1–52.

Enk JM, Devault AM, Kuch M, Murgha YE, Rouillard J-M, Poinar HN. 2014. Ancient Whole Genome Enrichment Using Baits Built from Modern DNA. Mol. Biol. Evol. 31:1292–1294.

Fabre P-H, Galewski T, Tilak M, Douzery EJP. 2013. Diversification of South American spiny rats (Echimyidae): a multigene phylogenetic approach: A multigene molecular phylogeny of South American spiny rats. Zool. Scr. 42:117–134.

Fabre P-H, Patton JL, Leite YL. 2016. Family Echimyidae (Hutias, Coypu, and South American Spiny rats). In: Handbook of the Mammals of the World. Vol. 6 - Lagomorphs and Rodents I. Lynx. Wilson DE, Lacher TE, Mittermeier RA. p. 552–641.

Fabre P-H, Upham NS, Emmons LH, Justy F, Leite YLR, Loss AC, Orlando L, Tilak M-K, Patterson BD, Douzery EJP. 2017. Mitogenomic phylogeny, diversification, and biogeography of South American spiny rats. Mol. Biol. Evol. 34:613–633.

Fabre P-H, Vilstrup JT, Raghavan M, Der Sarkissian C, Willerslev E, Douzery EJP, Orlando L. 2014. Rodents of the Caribbean: origin and diversification of hutias unravelled by next-generation museomics. Biol. Lett. 10:20140266–20140266.

Faircloth BC, McCormack JE, Crawford NG, Harvey MG, Brumfield RT, Glenn TC. 2012. Ultraconserved Elements Anchor Thousands of Genetic Markers Spanning Multiple Evolutionary Timescales. Syst. Biol. 61:717–726.

FitzJohn RG. 2012. Diversitree: comparative phylogenetic analyses of diversification in R. Methods Ecol Evol 3:1084-1092

Galewski T, Mauffrey J-F, Leite YLR, Patton JL, Douzery EJP. 2005. Ecomorphological diversification among South American spiny rats (Rodentia; Echimyidae): a phylogenetic and chronological approach. Mol. Phylogenet. Evol. 34:601–615.

Galtier N, Piganeau G, Mouchiroud D, Duret L. 2001. GC-content evolution in mammalian genomes: the biased gene conversion hypothesis. Genetics 159:907.

Giarla TC, Esselstyn JA. 2015. The Challenges of Resolving a Rapid, Recent Radiation: Empirical and Simulated Phylogenomics of Philippine Shrews. Syst. Biol. 64:727–740.

Gilbert PS, Chang J, Pan C, Sobel EM, Sinsheimer JS, Faircloth BC, Alfaro ME. 2015. Genome-wide ultraconserved elements exhibit higher phylogenetic informativeness than traditional gene markers in percomorph fishes. Mol. Phylogenet. Evol. 92:140–146.

Goodwin S, McPherson JD, McCombie WR. 2016. Coming of age: ten years of next-generation sequencing technologies. Nat. Rev. Genet. 17:333–351.

Hobolth A, Dutheil JY, Hawks J, Schierup MH, Mailund T. 2011. Incomplete lineage sorting patterns among human, chimpanzee, and orangutan suggest recent orangutan speciation and widespread selection. Genome Res. 21:349–356.

Huang H, Sukumaran J, Smith SA, Knowles LL. 2017. Cause of gene tree discord? Distinguishing incomplete lineage sorting and lateral gene transfer in phylogenetics. PeerJ Prepr. 5:e3489v1.

Huchon D, Douzery EJP. 2001. From the Old World to the New World: A Molecular Chronicle of the Phylogeny and Biogeography of Hystricognath Rodents. Mol. Phylogenet. Evol. 20:238–251.

Huchon D, Madsen O, Sibbald MJJB, Ament K, Stanhope MJ, Catzeflis F, Jong D, W W, Douzery EJP. 2002. Rodent Phylogeny and a Timescale for the Evolution of Glires: Evidence from an Extensive Taxon Sampling Using Three Nuclear Genes. Mol. Biol. Evol. 19:1053–1065.

Hugall AF, O'Hara TD, Hunjan S, Nilsen R, Moussalli A. 2016. An Exon-Capture System for the Entire Class Ophiuroidea. Mol. Biol. Evol. 33:281–294.

Husson AM. 1978. The mammals of Surinam. EJ Brill, Leiden, The Netherlands.

Källersjö M, Albert VA, Farris JS. 1999. Homoplasy Increases Phylogenetic Structure. Cladistics 15:91–93.

Katzman S, Kern AD, Bejerano G, Fewell G, Fulton L, Wilson RK, Salama SR, Haussler D. 2007. Human Genome Ultraconserved Elements Are Ultraselected. Science 317:915–915.

Kay RF, Madden RD, Cifelli RL, Flynn JJ. 1997. Vertebrate Paleontology in the Neotropics: The Miocene Fauna of La Venta, Colombia.

Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S, Buxton S, Cooper A, Markowitz S, Duran C, et al. 2012. Geneious Basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics 28:1647–1649.

de Knijff P. 2000. Messages through Bottlenecks: On the Combined Use of Slow and Fast Evolving Polymorphic Markers on the Human Y Chromosome. Am. J. Hum. Genet. 67:1055–1061.

Kozarewa I, Ning Z, Quail MA, Sanders MJ, Berriman M, Turner DJ. 2009. Amplification-free Illumina sequencing-library preparation facilitates improved mapping and assembly of (G+C)-biased genomes. Nat. Methods 6:291–295.

Kubatko LS, Degnan JH. 2007. Inconsistency of Phylogenetic Estimates from Concatenated Data under Coalescence. Syst. Biol. 56:17–24.

Kuraku S. 2013. Impact of asymmetric gene repertoire between cyclostomes and gnathostomes. Semin. Cell Dev. Biol. 24:119–127.

Lara MC, Patton JL, da Silva MNF. 1996. The Simultaneous Diversification of South American Echimyid Rodents (Hystricognathi) Based on Complete Cytochrome b Sequences. Mol. Phylogenet. Evol. 5:403–413.

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

Lartillot N, Philippe H. 2004. A Bayesian Mixture Model for Across-Site Heterogeneities in the Amino-Acid Replacement Process. Mol. Biol. Evol. 21:1095–1109.

Leite YL, Patton JL. 2002. Evolution of South American spiny rats (Rodentia, Echimyidae): the star-phylogeny hypothesis revisited. Mol. Phylogenet. Evol. 25:455–464.

Li C, Hofreiter M, Straube N, Corrigan S, Naylor GJP. 2013. Capturing protein-coding genes across highly divergent species. BioTechniques 54:321–326.

Loss AC, Moura RT, Leite YLR. 2014. Unexpected phylogenetic relationships of the painted tree rat *Callistomys pictus* (Rodentia: Echimyidae). Nat. Line 12:132–136.

Maddison WP, Wiens JJ. 1997. Gene Trees in Species Trees. Syst. Biol. 46:523–536.

Manthey JD, Campillo LC, Burns KJ, Moyle RG. 2016. Comparison of Target-Capture and Restriction-Site Associated DNA Sequencing for Phylogenomics: A Test in Cardinalid Tanagers (Aves, Genus: *Piranga*). Syst. Biol. 65:640–650.

McCartney-Melstad E, Mount GG, Shaffer HB. 2016. Exon capture optimization in amphibians with large genomes. Mol. Ecol. Resour. 16:1084–1094.

McCormack JE, Tsai WLE, Faircloth BC. 2016. Sequence capture of ultraconserved elements from bird museum specimens. Mol. Ecol. Resour. 16:1189–1203.

McKenna MC, Bell SK. 1997. Classification of mammals: above the species level. Columbia University Press

Meredith RW, Janečka JE, Gatesy J, Ryder OA, Fisher CA, Teeling EC, Goodbla A, Eizirik E, Simão TLL, Stadler T, et al. 2011. Impacts of the Cretaceous Terrestrial Revolution and KPg Extinction on Mammal Diversification. Science 334:521–524.

Mirarab S, Warnow T. 2015. ASTRAL-II: coalescent-based species tree estimation with many hundreds of taxa and thousands of genes. Bioinformatics 31:i44–i52.

Misof B, Liu S, Meusemann K, Peters RS, Donath A, Mayer C, Frandsen PB, Ware J, Flouri T, Beutel RG, et al. 2014. Phylogenomics resolves the timing and pattern of insect evolution. Science 346:763–767.

Moen D, Morlon H, Wiens JJ. 2015. Testing convergence Versus History: Convergence Dominates Phenotypic Evolution for over 150 Million Years in Frogs. Syst. Biol. 65:146-160

Nakamura K, Oshima T, Morimoto T, Ikeda S, Yoshikawa H, Shiwa Y, Ishikawa S, Linak MC, Hirai A, Takahashi H, et al. 2011. Sequence-specific error profile of Illumina sequencers. Nucleic Acids Res. 39:e90–e90.

Newman CE, Austin CC. 2016. Sequence capture and next-generation sequencing of ultraconserved elements in a large-genome salamander. Mol. Ecol. 25:6162–6174.

Nosenko T, Schreiber F, Adamska M, Adamski M, Eitel M, Hammel J, Maldonado M, Müller WEG, Nickel M, Schierwater B, et al. 2013. Deep metazoan phylogeny: When different genes tell different stories. Mol. Phylogenet. Evol. 67:223–233.

Nowak RM, Walker EP. 1999. Walker's Mammals of the World. JHU Press

Olivares AI, Verzi DH, Vucetich MG, Montalvo CI. 2012. Phylogenetic affinities of the late Miocene echimyid †*Pampamys* and the age of *Thrichomys* (Rodentia, Hystricognathi). J. Mammal. 93:76–86.

Pagel M. 1994. Detecting correlated evolution on phylogenies: a general method for the comparative analysis of discrete characters. Proc. R. Soc. Lond. B. 255:37-45

Peterson BK, Weber JN, Kay EH, Fisher HS, Hoekstra HE. 2012. Double Digest RADseq: An Inexpensive Method for De Novo SNP Discovery and Genotyping in Model and Non-Model Species. PLOS ONE 7:e37135.

Philippe H, Brinkmann H, Lavrov DV, Littlewood DTJ, Manuel M, Wörheide G, Baurain D. 2011. Resolving Difficult Phylogenetic Questions: Why More Sequences Are Not Enough. PLOS Biol. 9:e1000602.

Philippe H, Laurent J. 1998. How good are deep phylogenetic trees? Curr. Opin. Genet. Dev. 8:616–623.

R. Core Team. 2018. R: A language and environment for statistical computing. R Foundation for Statistical Computing. Vienna, Austria: ISBN 3-900051-07-0 Available from: https://www.R-project.org/

Ranwez V, Delsuc F, Ranwez S, Belkhir K, Tilak M-K, Douzery EJ. 2007. OrthoMaM: A database of orthologous genomic markers for placental mammal phylogenetics. BMC Evol. Biol. 7:241.

Ranwez V, Harispe S, Delsuc F, Douzery EJP. 2011. MACSE: Multiple Alignment of Coding SEquences Accounting for Frameshifts and Stop Codons. PLOS ONE 6:e22594.

Rodríguez F, Oliver JL, Marín A, Medina JR. 1990. The general stochastic model of nucleotide substitution. J. Theor. Biol. 142:485–501.

Rodríguez-Ezpeleta N, Brinkmann H, Roure B, Lartillot N, Lang BF, Philippe H, Anderson F. 2007. Detecting and Overcoming Systematic Errors in Genome-Scale Phylogenies. Syst. Biol. 56:389–399.

Rohland N, Malaspinas A-S, Pollack JL, Slatkin M, Matheus P, Hofreiter M. 2007. Proboscidean Mitogenomics: Chronology and Mode of Elephant Evolution Using Mastodon as Outgroup. PLOS Biol 5:e207.

Romiguier J, Ranwez V, Delsuc F, Galtier N, Douzery EJP. 2013. Less Is More in Mammalian Phylogenomics: AT-Rich Genes Minimize Tree Conflicts and Unravel the Root of Placental Mammals. Mol. Biol. Evol. 30:2134–2144.

Romiguier J, Roux C. 2017. Analytical Biases Associated with GC-Content in Molecular Evolution. Front. Genet. 8:16.

Schweizer RM, Robinson J, Harrigan R, Silva P, Galverni M, Musiani M, Green RE, Novembre J, Wayne RK. 2016. Targeted capture and resequencing of 1040 genes reveal environmentally driven functional variation in grey wolves. Mol. Ecol. 25:357–379.

Scornavacca C, Galtier N. 2017. Incomplete lineage sorting in mammalian phylogenetics. Syst. Biol. 66:112-120.

Simmons MP, Sloan DB, Springer MS, Gatesy J. 2019. Gene-wise resampling outperforms site-wise resampling in phylogenetic coalescence analyses. Mol. Phyl. Evol. 131:80-92.

Smith BT, Harvey MG, Faircloth BC, Glenn TC, Brumfield RT. 2014. Target Capture and Massively Parallel Sequencing of Ultraconserved Elements for Comparative Studies at Shallow Evolutionary Time Scales. Syst. Biol. 63:83–95.

Smith CH. 1842. Introduction to Mammals. In: The naturalist's library. Vol. 12. Chatto and Windus. London: W. Jardine.

Song S, Liu L, Edwards SV, Wu S. 2012. Resolving conflict in eutherian mammal phylogeny using phylogenomics and the multispecies coalescent model. Proc. Natl. Acad. Sci. U. S. A. 109:14942.

Soria-Carrasco V, Talavera G, Igea J, Castresana J. 2007. The K tree score: quantification of differences in the relative branch length and topology of phylogenetic trees. Bioinformatics 23:2954–2956.

Springer MS, DeBry RW, Douady C, Amrine HM, Madsen O, Jong D, W W, Stanhope MJ. 2001. Mitochondrial Versus Nuclear Gene Sequences in Deep-Level Mammalian Phylogeny Reconstruction. Mol. Biol. Evol. 18:132–143.

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

Steppan SJ, Adkins RM, Anderson J, Thorne J. 2004. Phylogeny and Divergence-Date Estimates of Rapid Radiations in Muroid Rodents Based on Multiple Nuclear Genes. Syst. Biol. 53:533–553.

Suren H, Hodgins KA, Yeaman S, Nurkowski KA, Smets P, Rieseberg LH, Aitken SN, Holliday JA. 2016. Exome capture from the spruce and pine giga-genomes. Mol. Ecol. Resour. 16:1136–1146.

Swofford DL. 2002. PAUP*: phylogenetic analysis using parsimony (* and other methods). Sunderland, MA.

Teasdale LC, Köhler F, Murray KD, O'Hara T, Moussalli A. 2016. Identification and qualification of 500 nuclear, single-copy, orthologous genes for the Eupulmonata (Gastropoda) using transcriptome sequencing and exon capture. Mol. Ecol. Resour. 16:1107–1123.

Thorne JL, Kishino H. 2002. Divergence Time and Evolutionary Rate Estimation with Multilocus Data.Neilsen R, editor. Syst. Biol. 51:689–702.

Thorne JL, Kishino H, Painter IS. 1998. Estimating the rate of evolution of the rate of molecular evolution. Mol. Biol. Evol. 15:1647–1657.

Tilak M-K, Justy F, Debiais-Thibaud M, Botero-Castro F, Delsuc F, Douzery EJP. 2014. A cost-effective straightforward protocol for shotgun Illumina libraries designed to assemble complete mitogenomes from non-model species. Conserv. Genet. Resour. 7:37–40.

Townsend JP, Naylor G. 2007. Profiling Phylogenetic Informativeness. Syst. Biol. 56:222–231.

Townsend JP, Su Z, Tekle YI. 2012. Phylogenetic signal and noise: predicting the power of a data set to resolve phylogeny. Syst. Biol. 61:835–849.

Upham NS, Patterson BD. 2012. Diversification and biogeography of the Neotropical caviomorph lineage Octodontoidea (Rodentia: Hystricognathi). Mol. Phylogenet. Evol. 63:417–429.

Upham NS, Patterson BD. 2015. Evolution of caviomorph rodents: a complete phylogeny and timetree for living genera. Biol. Caviomorph Rodents Divers. Evol. B. Aires SAREM Ser. A 1:63–120.

Van der Geer A, Lyras G, De Vos J, Dermitzakis M. 2011. Evolution of island mammals: adaptation and extinction of placental mammals on islands. John Wiley & Sons

Verzi DH. 2008. Phylogeny and adaptive diversity of rodents of the family Ctenomyidae (Caviomorpha): delimiting lineages and genera in the fossil record. J. Zool. 274:386–394.

Verzi DH, Olivares AI, Morgan CC. 2013. Phylogeny and Evolutionary Patterns of South American Octodontoid Rodents. Acta Palaeontol. Pol. 59:757–769.

Viswanathan VK, Krcmarik K, Cianciotto NP. 1999. Template secondary structure promotes polymerase jumping during PCR amplification. BioTechniques 27:508–511.

Vucetich MG, Mazzoni MM, Pardiñas UF. 1993. Los roedores de la Formación Collón Cura (Mioceno Medio) y la Ignimbrita Pilcaniyeu. Cañadón del Tordillo, Neuquén. Ameghiniana 30:361–381.

Vucetich MG, Verzi DH. 1991. Un nuevo Echimyidae (Rodentia, Hystricognathi) de la Edad Colhuehuapense de Patagonia y consideraciones sobre la sistemática de la familia. Ameghiniana 28:67–74.

Woods CA, Borroto Paéz R, Kilpatrick CW. 2001. Insular patterns and radiations of West Indian rodents. Biogeogr. West Indies Patterns Perspect.:335–353.

Woods CA, Howland EB. 1979. Adaptive Radiation of Capromyid Rodents: Anatomy of the Masticatory Apparatus. J. Mammal. 60:95–116.

Yang Y, Muzny DM, Reid JG, Bainbridge MN, Willis A, Ward PA, Braxton A, Beuten J, Xia F, Niu Z, et al. 2013. Clinical Whole-Exome Sequencing for the Diagnosis of Mendelian Disorders. N. Engl. J. Med. 369:1502–1511.

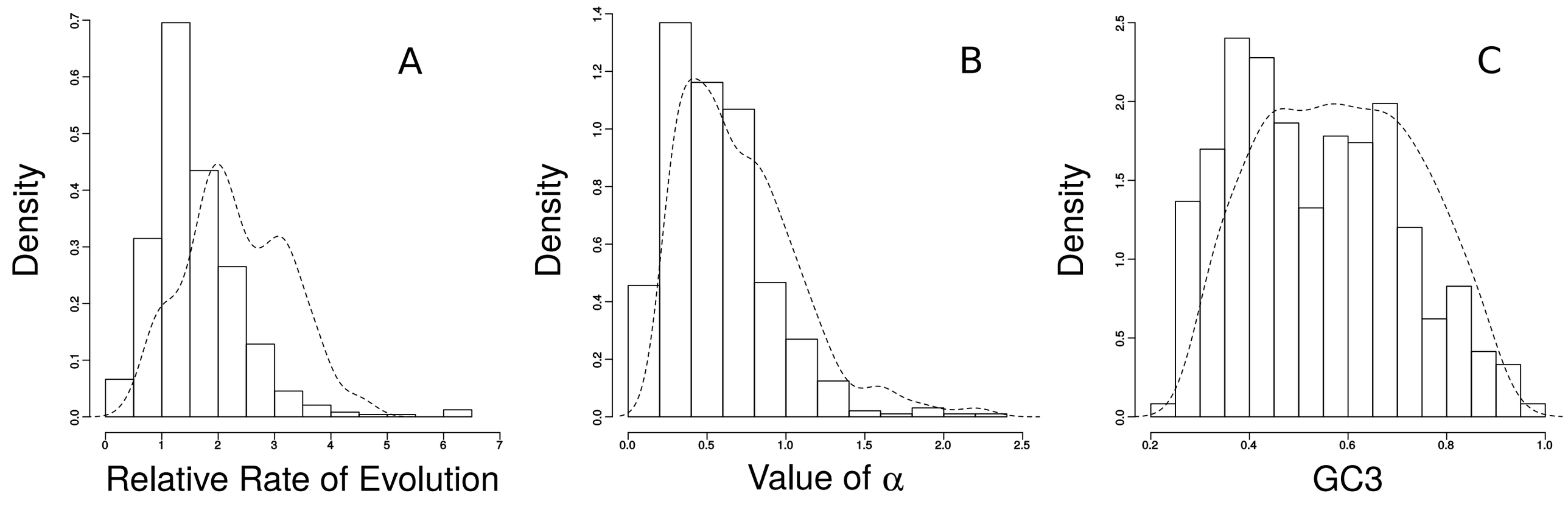
Yang Z. 1996. Among-site rate variation and its impact on phylogenetic analyses. Trends Ecol. Evol. 11:367–372.

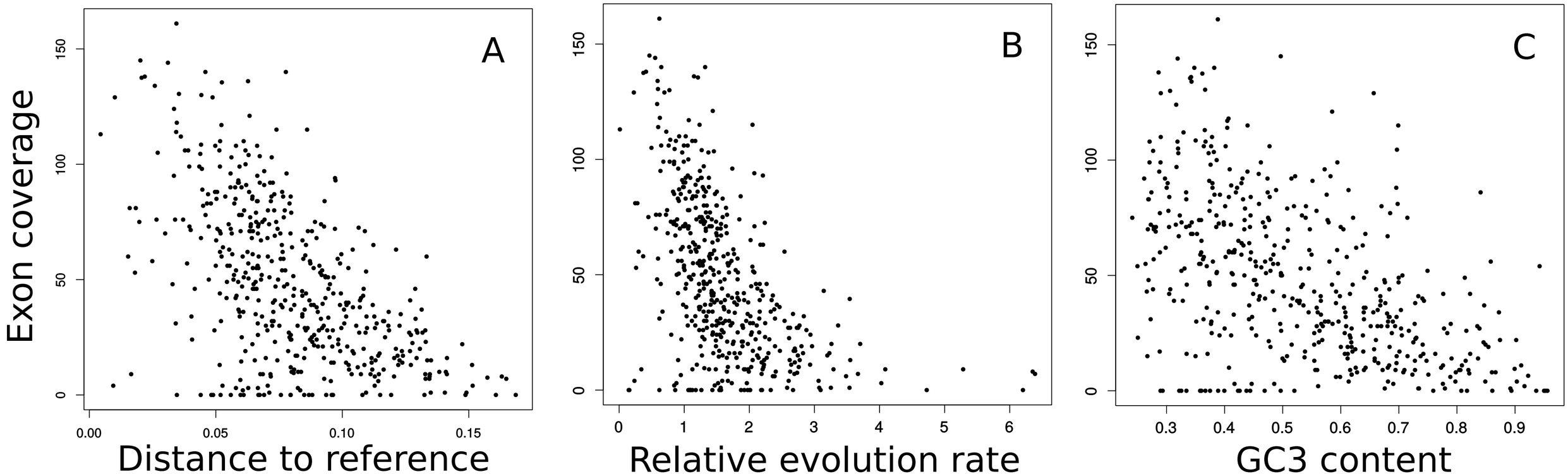
Yang Z, Rannala B. 2006. Bayesian Estimation of Species Divergence Times Under a Molecular Clock Using Multiple Fossil Calibrations with Soft Bounds. Mol. Biol. Evol. 23:212–226.

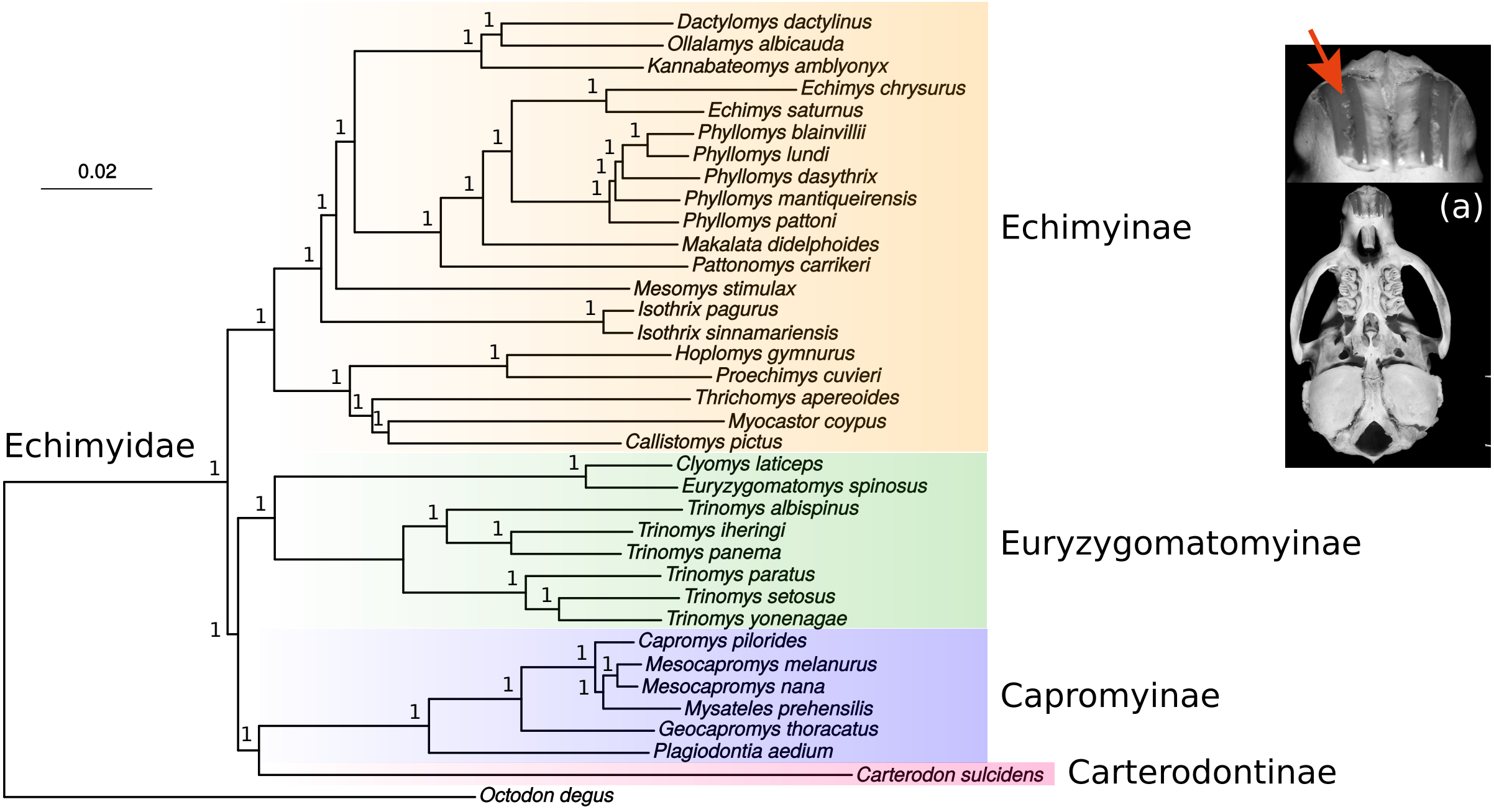
Yi X, Liang Y, Huerta-Sanchez E, Jin X, Cuo ZXP, Pool JE, Xu X, Jiang H, Vinckenbosch N, Korneliussen TS, et al. 2010. Sequencing of 50 Human Exomes Reveals Adaptation to High Altitude. Science 329:75–78.

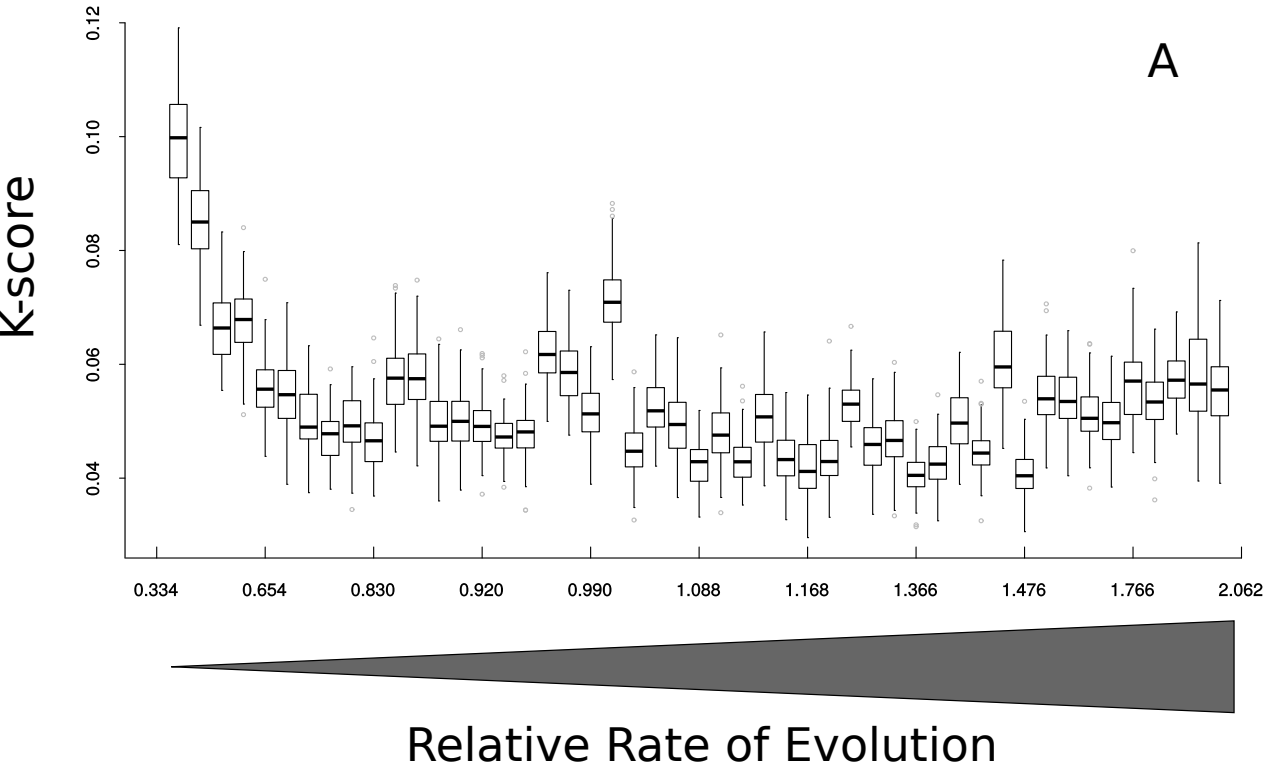
Zhang J, Chiodini R, Badr A, Zhang G. 2011. The impact of next-generation sequencing on genomics. J. Genet. Genomics 38:95–109.

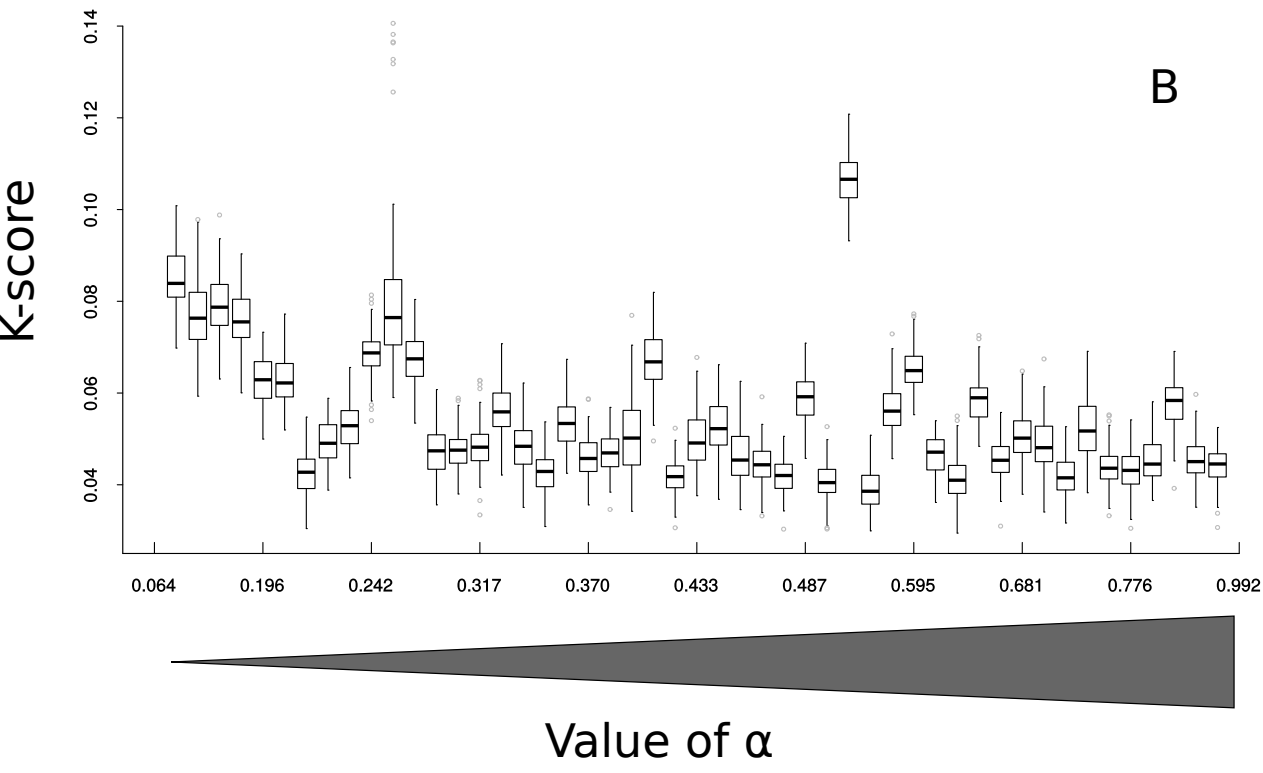
Zhang C, Rabiee M, Sayyari E, Mirarab S. 2018. ASTRAL-III: polynomial time species tree reconstruction from partially resolved gene trees. BMC Bioinformatics 19:153.

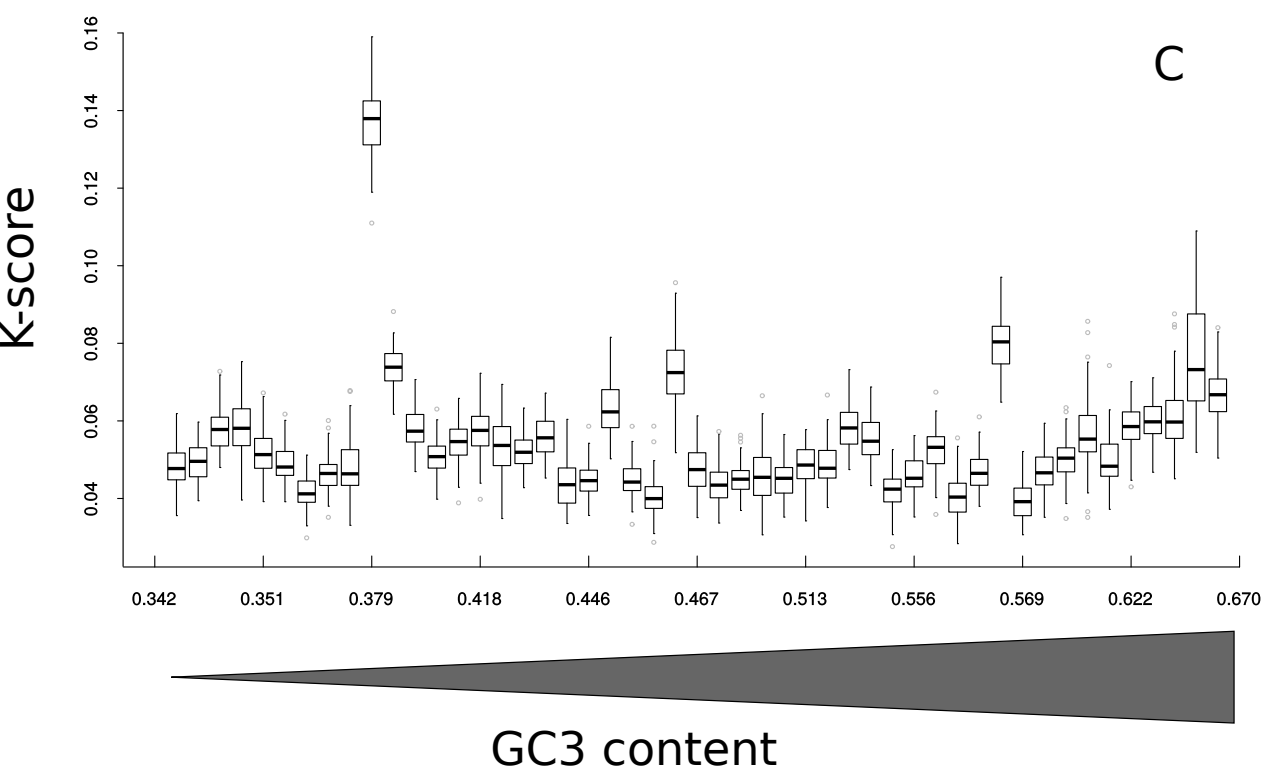

A
Density
Relative Rate of Evolution
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0
1
2
3
4
5
6
7
B
Density
Value of α
0.0
0.2
0.4
0.6
0.8
1.0
1.2
1.4
0.0
0.5
1.0
1.5
2.0
2.5
C
Density
GC3
0.0
0.5
1.0
1.5
2.0
2.5
0.2
0.4
0.6
0.8
1.0

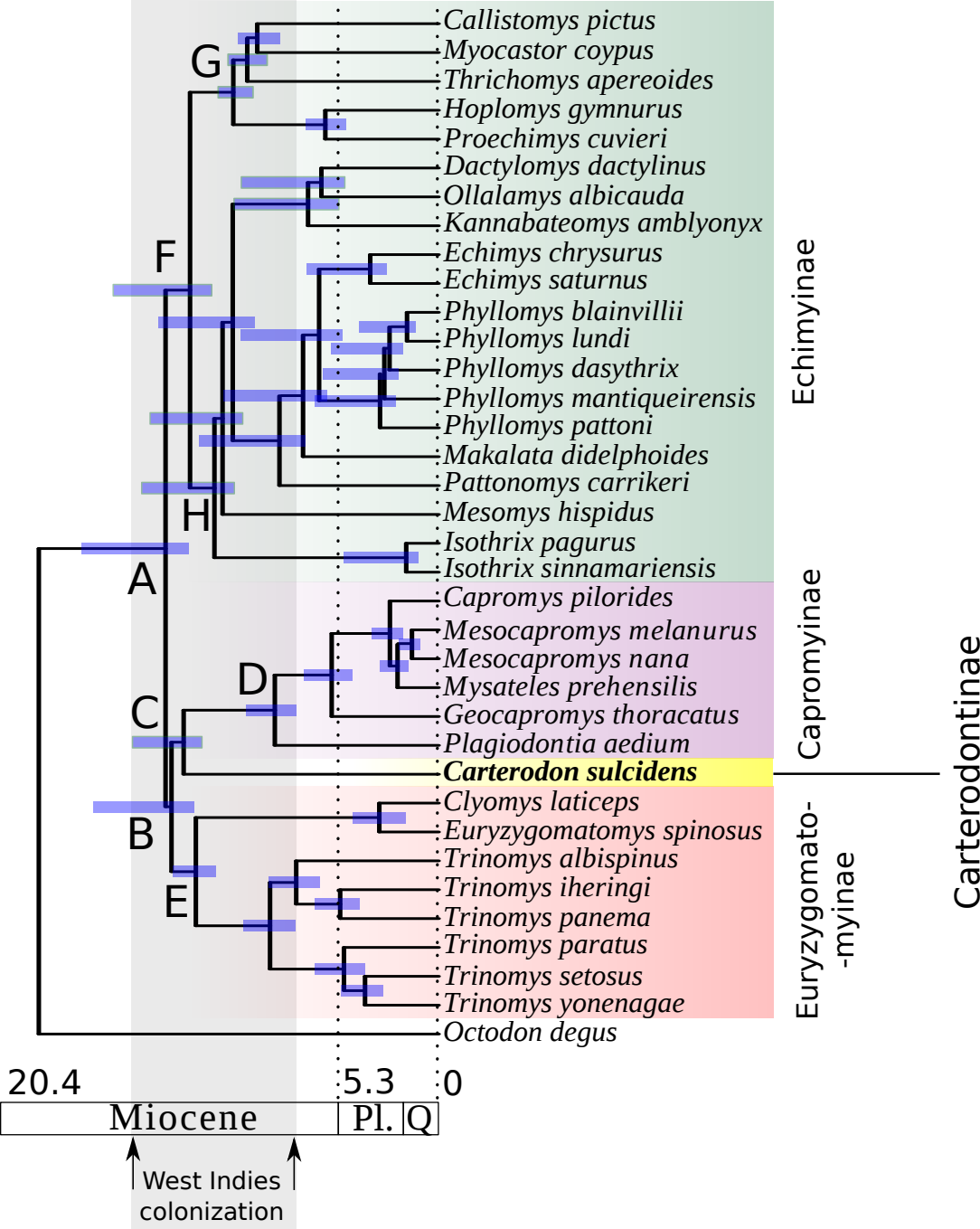

Exon coverage
A
B
C
0
50
100
150
0.00
0.05
0.10
0.15
Distance to reference
0
1
2
3
4
5
6
Relative evolution rate
0.3
0.4
0.5
0.6
0.7
0.8
0.9
GC3 content


0.02
Dactylomys dactylinus
Ollalamys albicauda
Kannabateomys amblyonyx
Echimys chrysurus
Echimys saturnus
Phyllomys blainvillii
Phyllomys lundi
Phyllomys dasythrix
Phyllomys mantiqueirensis
Phyllomys pattoni
Makalata didelphoides
Pattonomys carrikeri
Mesomys stimulax
Isothrix pagurus
Isothrix sinnamariensis
Hoplomys gymnurus
Proechimys cuvieri
Thrichomys apereoides
Myocastor coypus
Callistomys pictus
Clyomys laticeps
Euryzygomatomys spinosus
Trinomys albispinus
Trinomys iheringi
Trinomys panema
Trinomys paratus
Trinomys setosus
Trinomys yonenagae
Capromys pilorides
Mesocapromys melanurus
Mesocapromys nana
Mysateles prehensilis
Geocapromys thoracatus
Plagiodontia aedium
Carterodon sulcidens
Octodon degus
Echimyidae
Echimyinae
Euryzygomatomyinae
Capromyinae
Carterodontinae
(a)

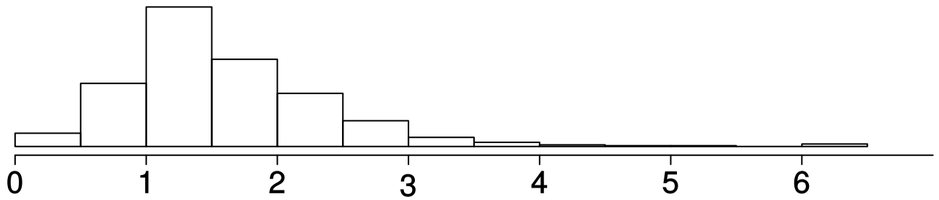

A
K-score
0.04
0.06
0.08
0.10
0.12
0.334
0.654
0.830
0.920
0.990
1.088
1.168
1.366
1.476
1.766
2.062
Relative Rate of Evolution


B
K-score
0.04
0.06
0.08
0.10
0.12
0.14
0.064
0.196
0.242
0.317
0.370
0.433
0.487
0.595
0.681
0.776
0.992
Value of α

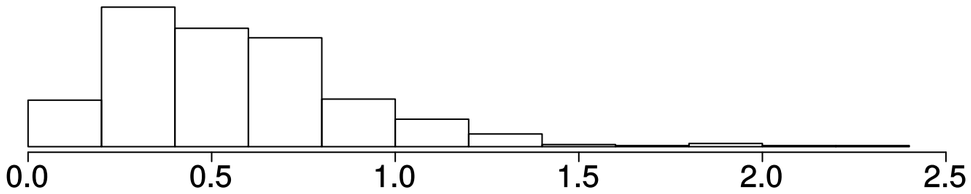

C
K-score
0.04
0.06
0.08
0.10
0.12
0.14
0.16
0.342
0.351
0.379
0.418
0.446
0.467
0.513
0.556
0.569
0.622
0.670
GC3 content


Callistomys pictus
Myocastor coypus
Thrichomys apereoides
Hoplomys gymnurus
Proechimys cuvieri
Dactylomys dactylinus
Ollalamys albicauda
Kannabateomys amblyonyx
Echimys chrysurus
Echimys saturnus
Phyllomys blainvillii
Phyllomys lundi
Phyllomys dasythrix
Phyllomys mantiqueirensis
Phyllomys pattoni
Makalata didelphoides
Pattonomys carrikeri
Mesomys hispidus
Isothrix pagurus
Isothrix sinnamariensis
Capromys pilorides
Mesocapromys melanurus
Mesocapromys nana
Mysateles prehensilis
Geocapromys thoracatus
Plagiodontia aedium
Carterodon sulcidens
Clyomys laticeps
Euryzygomatomys spinosus
Trinomys albispinus
Trinomys iheringi
Trinomys panema
Trinomys paratus
Trinomys setosus
Trinomys yonenagae
Octodon degus
A
B
C
D
E
F
G
H
Echimyinae
Capromyinae
Euryzygomato-
-myinae
Carterodontinae
20.4
5.3
0
Miocene
Pl.
Q
West Indies
colonization

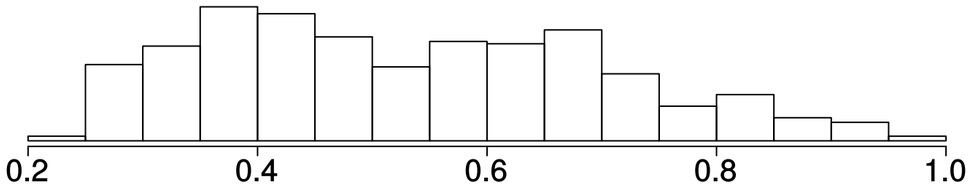
483 nuclear DNA exons
with contrasting evolutionary properties



Relative Rate of Evolution



Substitution heterogeneity



G+C on third codon position

Schematic phylogenetic tree
of South American spiny rats