Molecular Phylogenetics and Evolution 175 (2022) 107576

Contents lists available at ScienceDirect

Molecular Phylogenetics and Evolution

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

Old specimens for old branches: Assessing effects of sample age in resolving a rapid Neotropical radiation of squirrels

Edson F. Abreu [a,b,*], Silvia E. Pavan [b,1], Mirian T.N. Tsuchiya [b,c], Bryan S. McLean [d], Don E. Wilson [e], Alexandre R. Percequillo [a], Jesús E. Maldonado [b]

[a] Laboratório de Mamíferos, Departamento de Ciências Biológicas, Escola Superior de Agricultura Luiz de Queiroz, Universidade de São Paulo, Piracicaba, SP, Brazil
[b] Center for Conservation Genomics, Smithsonian National Zoo and Conservation Biology Institute, Washington, DC, USA
[c] Data Science Lab, Office of the Chief Information Officer, Smithsonian Institution, Washington, DC, USA
[d] Department of Biology, University of North Carolina Greensboro, Greensboro, NC, USA
[e] Division of Mammals, National Museum of Natural History, Smithsonian Institution, Washington, DC, USA

ARTICLE INFO

Keywords:
Deep sequencing
Filtering approaches
Historical samples
Phylogenomics
Sciuridae
Ultraconserved Elements

ABSTRACT

Ultraconserved Elements (UCEs) have been useful to resolve challenging phylogenies of non-model clades, unpuzzling long-conflicted relationships in key branches of the Tree of Life at both deep and shallow levels. UCEs are often reliably recovered from historical samples, unlocking a vast number of preserved natural history specimens for analysis. However, the extent to which sample age and preservation method impact UCE recovery as well as downstream inferences remains unclear. Furthermore, there is an ongoing debate on how to curate, filter, and properly analyze UCE data when locus recovery is uneven across sample age and quality. In the present study we address these questions with an empirical dataset composed of over 3800 UCE loci from 219 historical and modern samples of Sciuridae, a globally distributed and ecologically important family of rodents. We provide a genome-scale phylogeny of two squirrel subfamilies (Sciurillinae and Sciurinae: Sciurini) and investigate their placement within Sciuridae. For historical specimens, recovery of UCE loci and mean length per locus were inversely related to sample age; deeper sequencing improved the number of UCE loci recovered but not locus length. Most of our phylogenetic inferences—performed on six datasets with alternative data-filtering strategies, and using three distinct optimality criteria—resulted in distinct topologies. Datasets containing more loci (40% and 50% taxa representativeness matrices) yielded more concordant topologies and higher support values than strictly filtered datasets (60% matrices) particularly with IQ-Tree and SVDquartets, while filtering based on information content provided better topological resolution for inferences with the coalescent gene-tree based approach in ASTRAL-III. We resolved deep relationships in Sciuridae (including among the five currently recognized subfamilies) and relationships among the deepest branches of Sciurini, but conflicting relationships remain at both genus- and species-levels for the rapid Neotropical tree squirrel radiation. Our results suggest that phylogenomic consensus can be difficult and heavily influenced by the age of available samples and the filtering steps used to optimize dataset properties.

1. Introduction

Targeted capture of Ultraconserved Elements (UCEs) is a powerful technique to obtain thousands of loci across the nuclear genome for phylogenomic inference (Faircloth et al., 2012). UCEs have been useful to resolve especially challenging phylogenies, unpuzzling long-conflicted relationships in the Tree of Life at both deep and shallow levels (Baca et al., 2017; Crawford et al., 2012; Faircloth et al., 2013; Oliveros et al., 2019; Streicher and Wiens, 2017), including the most diverse radiation of mammals, the order Rodentia (Esselstyn et al., 2017; Hawkins et al., 2016; Parada et al., 2021; Swanson et al., 2019). One great advantage of UCEs for phylogenetic studies is that they contain—besides the highly conserved core regions useful for inference in deep time—flanking regions that are sufficiently variable among closely related taxa (Faircloth et al., 2012). Therefore, UCEs can be informative for evolutionary radiations unfolding at non-constant rates over time

* Corresponding author at: Department of Mammalogy, American Museum of Natural History, New York, NY, USA.
E-mail address: edson.fd.abreu@gmail.com (E.F. Abreu).
[1] Present address: Department of Biological Sciences, California State Polytechnic University, Humboldt, Arcata, CA, USA.

https://doi.org/10.1016/j.ympev.2022.107576
Received 21 February 2022; Received in revised form 10 June 2022; Accepted 1 July 2022
Available online 6 July 2022
1055-7903/© 2022 Elsevier Inc. All rights reserved.

(McLean et al., 2019). In addition, because UCEs are often recoverable from historical samples, they make a vast number of the world's preserved natural history specimens available for analysis, for which few or no other comparable materials may be available (Branstetter et al., 2021; Castañeda-Rico et al., 2020; Derkarabetian et al., 2019; McCormack et al., 2016; McDonough et al., 2018; Ruane and Austin, 2017).

Despite these advantages, there is an ongoing debate on how to curate and properly analyze UCE data, especially as capture and sequencing success may vary with sample age (Card et al., 2021) and UCE-based inferences may be sensitive to taxon sampling strategy, data filtering options, and inference methods (Hosner et al., 2016; Platt et al., 2018; Streicher et al., 2016). Regarding data filtering, some approaches are available (see McLean et al., 2019; Platt et al., 2018; Simmons et al., 2016; Streicher et al., 2016) and two of the most common are: (1) sample completeness, or percentage of samples present in a given locus in the final matrix; and (2) information content, or the percentage of informative sites per UCE locus. Both approaches ultimately define the number of loci to be analyzed, and the choice of a filtering strategy (if any) and associated parameters may be especially critical for datasets dominated by degraded DNA from historical samples (Derkarabetian et al., 2019; McDonough et al., 2018). Increasing the number of loci in the datasets can maximize the robustness of the phylogenetic inferences in terms of both statistical support and topological consistency, although this is most often observed in concatenation-based inference (Dornburg et al., 2019; McLean et al., 2019; Platt et al., 2018). On the other hand, a filtering strategy based on the information content may strongly benefit gene-tree based coalescent methods by minimizing gene tree estimation error, as these methods are sensitive to imprecise gene trees provided by loci with low information content, which can support alternative topologies with equal probability (Blom et al., 2017; Hahn and Nakhleh, 2016; Mirarab et al., 2016). This approach may especially benefit inferences in rapid radiations, which—due to the presence of low phylogenetic signal—represent some of the most difficult phylogenetic problems to solve (McLean et al., 2019; Smith et al., 2014).

Two clades of squirrels (Family Sciuridae, containing about 300 species globally; Burgin et al., 2018) are found in the Neotropics. One of those is represented by the Neotropical pygmy squirrel, *Sciurillus pusillus*, which is included in the monotypic subfamily Sciurillinae (Thorington et al., 2012). This enigmatic squirrel diverged early in the history of the family, about 35 million years ago (Mya; Mercer and Roth, 2003; Upham et al., 2019), and its phylogenetic affinities remain unclear (Fabre et al., 2012; Mercer and Roth, 2003; Steppan et al., 2004; Upham et al., 2019; Zelditch et al., 2015). Its distribution suggests that squirrels have a longer evolutionary history in South America than is evident from the remainder of sciurid diversity. It is also noteworthy that no phylogenetic study published so far included more than two samples of *S. pusillus*, and consequently the genetic diversity within this *trans*-Amazonian taxon (Vivo and Carmignotto, 2015) is virtually unknown.

The other clade, the subfamily Sciurinae, tribe Sciurini, is a diverse radiation of Holarctic and Neotropical forest-dwelling squirrels that originated around 14 Mya, most likely in North America (Abreu-Jr et al., 2020a; Mercer and Roth, 2003; Thorington et al., 2012); but see Pečnerová et al. (2015) and Rocha et al. (2016) for alternative hypotheses. Tree squirrels exhibit strikingly high rates of diversification (Roth and Mercer, 2008), especially in the Neotropics, where they experienced rapid lineage splitting after the South American invasion, around 6 Mya (Abreu-Jr et al., 2020a). Until recently, the genetic diversity and molecular systematics of tree squirrels were poorly known, as the few published studies were very limited in terms of taxonomic coverage, particularly regarding Neotropical taxa, and were composed of a few mitochondrial and nuclear genes (e.g., Oshida et al., 2009; Pečnerová and Martínková, 2012; Villalobos and Gutierrez-Espeleta, 2014; Pečnerová et al., 2015; Aghbolaghi et al., 2019, 2020). Abreu-Jr et al. (2020b) published heretofore the most comprehensive phylogenetic study for tree squirrels, including mitochondrial genome data from 43 of 46 putative Sciurini species (sensu Thorington et al., 2012; Vivo and Carmignotto, 2015; Hope et al., 2016). Despite the phylogenetic and biogeographic advances established in these studies (see also Abreu-Jr et al., 2020a), further perspectives from independent nuclear loci are still missing.

In this study we provide a nuclear genome-scale perspective of the phylogeny of Neotropical squirrels (Sciurillinae and Sciurinae: Sciurini) and investigate the placement of these subfamilies within Sciuridae—a globally distributed and ecologically important family of rodents. We employed over 3800 UCE loci sequenced from 219 specimens, including substantial numbers of historical specimens with low-quality tissue preservation, a scenario common for many Neotropical species that are poorly sampled throughout their ranges. We therefore also leveraged this dataset to investigate how sample age (i.e., year of museum specimen collection) influenced the number of UCE loci recovered and the average length of those loci, and to examine whether deeper sequencing mitigates both issues for historical samples. We inferred phylogenies using three conceptually distinct optimality criteria [a concatenated method in IQ-Tree (Minh et al., 2020), a coalescent gene-tree based method in ASTRAL-III (Zhang et al., 2018), and a coalescent site-based method in SVDquartets (Chifman and Kubatko, 2014)] and gauged robustness of our inferences by varying two dataset properties (sample completeness per locus, and proportion of variable sites per locus) to verify how combinations of these distinct parameters affected phylogenetic consistency and support for still-dynamically changing taxon concepts for tree squirrels.

2. Material and methods

2.1. Sampling

We obtained frozen or ethanol-preserved tissues (hereafter modern samples) from 188 specimens and we sampled remains of muscular tissue adherent to skulls or skin clips (hereafter historical samples) from 94 dry museum specimens. Historical samples are crucial, as they represent testimonies of species not recently sampled and/or of critically important geographic variants, especially in the incompletely-delineated Neotropical taxa. Sampling from dry museum specimens followed rigorous procedures to avoid cross-contamination, as described in McDonough et al. (2018) and Abreu-Jr et al. (2020b). We also included UCE data available at GenBank for three specimens of Callosciurinae (Hawkins et al., 2016), totaling 285 samples analyzed in this study. Voucher material of all samples are housed in 27 scientific collections from South America, North America, and Europe (see Supplementary Appendix 1). To provide broadest possible phylogenetic context, samples include the only representative of Sciurillinae (*Sciurillus pusillus*), one species of Ratufinae (*Ratufa bicolor*), two samples of Xerinae (*Marmota monax*), three species of Callosciurinae (*Callosciurus adamsi*, *Exilisciurus exilis*, and *Lariscus insignis*), and 45 species of Sciurinae, two being from the tribe Pteromyini (*Glaucomys volans* and *Hylopetes phayrei*) and 43 from the tribe Sciurini. We also sampled two individuals of *Aplodontia rufa* (family Aplodontiidae, sister to Sciuridae; Upham et al., 2019) as an unambiguous outgroup. A complete list of specimens accompanied by catalogue and geographic data and other relevant information is provided in Supplementary Table S1.

2.2. DNA extraction and library preparation

DNA extractions of modern samples were performed using the DNeasy® Blood & Tissue kit, following manufacture's protocol (Qiagen Inc.). DNA of historical samples was extracted using a standard phenol–chloroform protocol (see detailed protocol in McDonough et al., 2018 and also additional information in Abreu-Jr et al., 2020b) in an isolated ancient DNA facility at the Smithsonian's Center for Conservation Genomics. Total DNA concentrations were measured using a Qubit 2.0 fluorometer (Thermo Fisher Scientific). DNA extracted from ethanol-preserved tissues was sheared with QSonica Q800R using 25 %

of amplitude and 5 min of on/off pulse. Sheared DNA was visualized on agarose gel to confirm the resulting fragment size around 300 bp. Approximately 500 ng of sheared modern DNA and about 35 μl of historical DNA (regardless of concentration) were purified using 5x SPRI magnetic beads (Rohland and Reich, 2012).

Library preparations were performed using the KAPA LTP Library Preparation Kit (Roche Sequencing) following the manufacturer's protocol. Nextera-style indices and KAPA HiFi Hotstart ReadyMix (Roche Sequencing) were used for indexing PCRs (iPCR). The iPCR profile followed Abreu-Jr et al. (2020b). Amplified samples were purified using 1.8x SPRI magnetic beads, quantified with a Qubit 2.0 fluorometer (Thermo Fisher Scientific) and visualized on a 1.5% agarose gel.

2.3. UCE capture and sequencing

Libraries were multiplexed in equimolar ratios for UCE enrichment. We pooled up to four libraries for historical samples and up to eight libraries for modern samples. No historical samples were pooled with modern samples to avoid biased enrichment. Multiplexed libraries were then dried out and re-diluted in 7 μl of nuclease-free water. UCE enrichments were performed using the myBaits UCE Tetrapods 5Kv1 kit (Arbor Biosciences), containing a probe set of around 5000 UCE loci. Hybridization temperatures were 60 °C for historical samples and 65 °C for modern samples. Post-capture amplifications were performed using KAPA HiFi Hotstart ReadyMix (Roche Sequencing), with the following profile: initial denaturation at 98 °C for 2 min, a final extension at 72 °C for 7 min, and 15 (for modern samples) or 16 (for historical samples) cycles of amplification, with denaturation at 98 °C for 20 sec, annealing at 60 °C for 30 sec, and extension at 72 °C for 30 sec. A 1.8x SPRI magnetic beads cleanup was performed subsequently.

Cleaned amplification products were quantified using a Qubit 2.0 fluorometer and visualized on a Bioanalyzer (Agilent) with high sensitivity kits. Equimolar pooling of samples was performed based on the concentration (ng/μl) and on the average size (bp) of amplified fragments. Due to the high concentration of dimers obtained, especially for historical samples, we size-selected fragments between 200 and 550 bp using a Pippin Prep (Sage Science). Samples were sequenced in four distinct Illumina runs: one including 63 historical samples on a Hi-Seq 2500 125 PE at the DNA Sequencing Center at Brigham Young University, Utah; two composed of 96 modern samples each, and the last composed of 86 historical samples (50 samples from this project and 36 from a different project), on a Hi-Seq 4000 150 PE at the Vincent J. Coates Genomics Sequencing Laboratory at the University of California, Berkeley. Sequencing reactions included Illumina Free Adapter Blocking Reagent to prevent index hopping.

2.4. Data processing and samples filtering

We followed the PHYLUCE 1.6 pipeline (Faircloth, 2016; Faircloth et al., 2012) adapted for the Smithsonian Institution High Performance Cluster (SI-HPC; https://doi.org/10.25572/SIHPC) to process UCE data (https://github.com/SmithsonianWorkshops/Targeted_Enrichment/). We cleaned adapter contamination and low-quality bases using Illumiprocessor 2.0 (Bolger et al., 2014; Faircloth, 2013). Reads were assembled into contigs with Trinity (Grabherr et al., 2011) and contigs were matched to the uce-5k-probe-set (Faircloth et al., 2012). To generate the initial alignments, we selected only samples for which at least 400 UCE loci were enriched (48 samples were dropped). This threshold was chosen based on preliminary phylogenetic analyses that indicated samples with less than around 400 consistently enriched loci clustered together (regardless of their taxonomic identification) and presented extremely long branch lengths in phylogenetic inferences. Those samples were mostly historical and had short mean locus lengths (see "Sequencing success" section in the Results). UCE loci were aligned using MAFFT 7 (Katoh and Standley, 2013; Nakamura et al., 2018) with most default settings but considering max-divergence = 0.15 and edge-trimming only. We did not use the default internal trimming routine of PHYLUCE based on Gblocks (Castresana, 2000; Talavera and Castresana, 2007) because locus alignments were unambiguous and recent studies have shown that Gblocks tends to produce biased results in the context of phylogenetic reconstructions (Portik and Wiens, 2021; Tan et al., 2015).

We concatenated all loci in which at least 50% of the samples were present (see more details on this approach in the next section) and trimmed the supermatrix with SPRUCEUP (Borowiec, 2019) to remove any poorly aligned sequences, which can occur in target enrichment studies containing old samples (Branstetter et al., 2021). We ran SPRUCEUP using the uncorrected p-distances, selected a window size of 50 bp with overlap size of 25 bp, and applied a cutoff of 0.98 over a lognormal distribution. These settings were chosen after several rounds of parameter testing. The combination of the distance model employed with the cutoff value of 0.98 provided the best outcome as it helped to reduce most of the extremely long branches without collapsing recent evolutionary lineages with shallow divergences.

We then used this trimmed matrix to perform a set of ML-based exploratory analyses (see below for methodological procedures) to identify potential nonorthologs among our samples. Nonorthologs can be detected by identifying highly divergent conspecific or congeneric sequences and by their characteristic exaggerated long branches (Derkarabetian et al., 2019). We identified and excluded 18 persistently problematic and low-quality samples from the downstream analyses, usually from old specimens, which did not show improvement on the exaggerated branch lengths even after been treated with SPRUCEUP. It is important to note that those samples removed at this step represented mostly additional specimens of species already present in our dataset.

2.5. UCE data filtering approaches

To generate the final data matrices for phylogeny inference, we selected the 219 samples that passed our quality tests and we applied two locus-based strategies of data filtering. Our main goal with this filtering strategy was to evaluate if inclusion of more samples per locus altered phylogenetic inferences relative to inclusion of fewer samples with lower rates of overall missing data. First, we generated matrices with distinct percentages of sample representativeness per UCE locus, i.e., the sample completeness. Allowing lower values of sample completeness (i.e., more missing samples per locus) causes the number of UCEs in the final data matrices to increase substantially, but with an overall increase of missing data. We generated three distinct datasets, each one including UCEs that were enriched in at least 60% (2069 UCEs), 50% (3545 UCEs), and 40% (3841 UCEs) of samples, using the "phyluce_align_get_only_loci_with_min_taxa" function (Faircloth, 2016). We attempted to also generate a 70% matrix but this value allowed us to retain only 43 UCEs, which we did not consider sufficient data to conduct robust multilocus inferences.

The datasets above were the basis for the second filtering strategy focusing on information content. We used the function "phyluce_align_get_informative_sites" to create three new datasets including only UCEs with 5% or more variable sites. This reduced the number of UCE loci in each matrix to 977 loci in the 60% sample representativeness matrix, 1742 loci in the 50% matrix, and 2112 loci in the 40% matrix (Table 1). The percentage of variable sites was quantified relatively to the trimmed locus length. According to McLean et al. (2019), using the proportion of variable sites as a metric to quantify information content and guide dataset filtering is potentially advantageous because it captures intrinsic properties of loci (e.g., substitution rate and numbers of parsimony informative sites) but does not require previous inferences of genes or species trees that could introduce additional errors. The step-by-step workflow of our data filtering approach is presented as Supplementary Fig. S1.

Table 1
Summary characteristics of the UCE datasets analyzed in this study.

	Matrix completeness	Filter for information content	Number of UCE loci included	Total base-pairs	Parsimony-informative sites	Overall missing data
Dataset 1	40%	No	3841	1,688,462	86,554	52.63%
Dataset 2	50%	No	3545	1,572,845	81,417	51.72%
Dataset 3	60%	No	2069	937,316	47,153	49.49%
Dataset 4	40%	≥5% of IS*	2112	927,942	72,468	52.94%
Dataset 5	50%	≥5% of IS	1742	763,974	62,663	51.82%
Dataset 6	60%	≥5% of IS	977	438,765	35,589	49.42%

* Informative Sites (IS).

2.6. UCE sequencing success

Given the large number of historical samples utilized, we also investigated whether the sample age influenced UCE data generation by exploring two metrics of capture/sequencing success: the number of UCE loci recovered per sample and the mean UCE length. These analyses were performed prior to any of the filtering steps above. To avoid potential bias associated to historical sample storage (see McDonough et al., 2018), we included in this comparison only samples from North American museums, which represent the great majority of our historical samples (80 out of 94 samples) and are stored with similar procedures and standards. However, our modern sample set included both North American (n = 96) and South American (n = 69) collections, as exploratory Mann-Whitney U Tests revealed no significant differences between these two groups in the number of UCE recovered ($p = 0.089$) nor in the mean length of UCE locus ($p = 0.076$). We performed linear regressions of each of the two metrics above on sample age using the R function "lm" from the package stats v3.6.2 (R Core Team, 2016). Moreover, we examined the impact of deeper sequencing for these two metrics (recovery of UCE loci, mean length of UCEs) by doubling the sequencing effort for a subset of 14 historical samples collected from 1900 to 1987 (via two different sequencing runs).

2.7. Phylogenetic inferences

We performed phylogenetic inferences using concatenation (Maximum Likelihood [ML] method) and two distinct coalescent approaches (a two-step summary coalescent method [species tree inferred from individual gene trees] and a site-based method). For the ML analyses, concatenated matrices were generated using the function "phyluce_align_format_nexus_files_for_raxml" (Faircloth, 2016), and analyses were performed in IQ-Tree 2.1.1 (Minh et al., 2020). IQ-Tree analyses were performed with the dataset partitioned by locus and with GTR + G4 model of substitution set for each locus. We selected this model for all loci without performing model testing because more complex models do not negatively impact phylogenetic estimations even if less complex models would be preferred for some loci (Abadi et al., 2019). An initial parsimony tree was created by phylogenetic likelihood library (PLL), and branch support was assessed from 1000 ultrafast bootstrap replicates (UFBoot2; Hoang et al., 2018). The final trees displayed were the bootstrap consensus trees. IQ-Tree analyses were performed in CIPRES (Miller et al., 2010).

We employed ASTRAL-III (Zhang et al., 2018) to perform the summary coalescent analyses. Individual gene trees were inferred in RAxML 8.2.7 (Stamatakis, 2014) by running ten independent searches under the GTR nucleotide substitution model with gamma-distributed rate heterogeneity (GTRGAMMA). The best-scoring ML trees were selected to draw the bootstrap support values obtained from 1000 replicates using the "thorough standard bootstrap" optimization option. All RAxML analyses were conducted at the SI-HPC (https://doi.org/10.25572/SIHPC). The gene trees with the highest likelihood score in each analysis were used as input trees in ASTRAL-III. Inferences in ASTRAL-III were performed with default parameters, and were run in two-ways, the first considering the specimens as terminals and the second including a mapping file to perform species trees.

Site-based coalescent analyses were performed using SVDquartets (Chifman and Kubatko, 2014) implemented in PAUP* 4.0a166 (Swofford, 2003). We ran exhaustive quartet sampling analyses—which evaluates all possible quartets—to search for the best tree for each dataset. As in ASTRAL-III, we first ran all analyses considering the specimens as terminals and subsequently we inferred species trees by setting up a taxa (species) partition. To assess support for each tree, we ran a new set of non-exhaustive analyses assessing 1,000,000 randomly sampled quartets and performing 100 standard bootstrap replicates. Nodal support was calculated from the bootstrap replicates and added to the exhaustive quartet sampling tree using the function "addConfidences" in the R package phangorn (Schliep, 2011).

2.8. Topological comparisons

We visualized overall nodal support and support for the supraspecific relationships obtained for the six distinct datasets (matrices with 40, 50, and 60% of completeness, and with or without filter for informative sites) within each inference method using boxplots created with the function "geom_boxplot" in the R package ggplot2 (Wickham, 2016). We quantified discordances among topologies—both within and among methods—based on the Robinson-Foulds (RF) distance metric, using the function "RF.dist" in the R package phangorn (Schliep, 2011). Pairwise RF distances were scaled to enable comparisons among methods. To visualize those differences, we performed pairwise comparisons of trees, creating heat maps with the function "levelplot" from lattice (Sarkar, 2008). Subsequently, we used the Metric Multidimensional Scaling (MDS, aka Principal Coordinates Analysis, PCoA) to summarize tree distances ("dudi.pco" function in adegenet; Jombart 2008) and visualize them in a two-dimensional space using ggplot2 (Wickham, 2016).

3. Results

3.1. Sequencing success and relationship with sample age

We sequenced 282 samples, 94 from historical museum specimens and 188 from modern specimens. The mean number of paired-end reads recovered for historical samples was 2,584,400 (range 32,012–15,776,174) and for modern samples was 3,344,948 (range 4077–12,726,177). On a per-sample basis, Trinity assemblies contained an average of 15,501 contigs per historical sample (range 8–226,333) and 41,704 contigs per modern sample (range 101–317,895). Assembled contigs were recovered from an average of 720 UCE loci per historical sample (range 0–2419) with mean length of 274 bp, and an average of 2459 UCE loci per modern sample (range 18–3385) with mean length of 454 bp (detailed assembly statistics are in Supplementary Table S2). Modern samples thus performed significantly better than historical samples both in number of UCE loci enriched per sample and UCE mean length ($W = 995$, $p < 0.001$, and $W = 867$, $p < 0.001$, respectively).

We observed that sample age was a significant predictor of the number of UCE loci recovered for historical samples ($R^2 = 0.1072$, $p = 0.0034$; Fig. 1A) and that the average UCE locus length was also positively correlated with sample age ($R^2 = 0.0544$, $p = 0.0419$; Fig. 1B); whereas for modern samples there was no correlation between sample

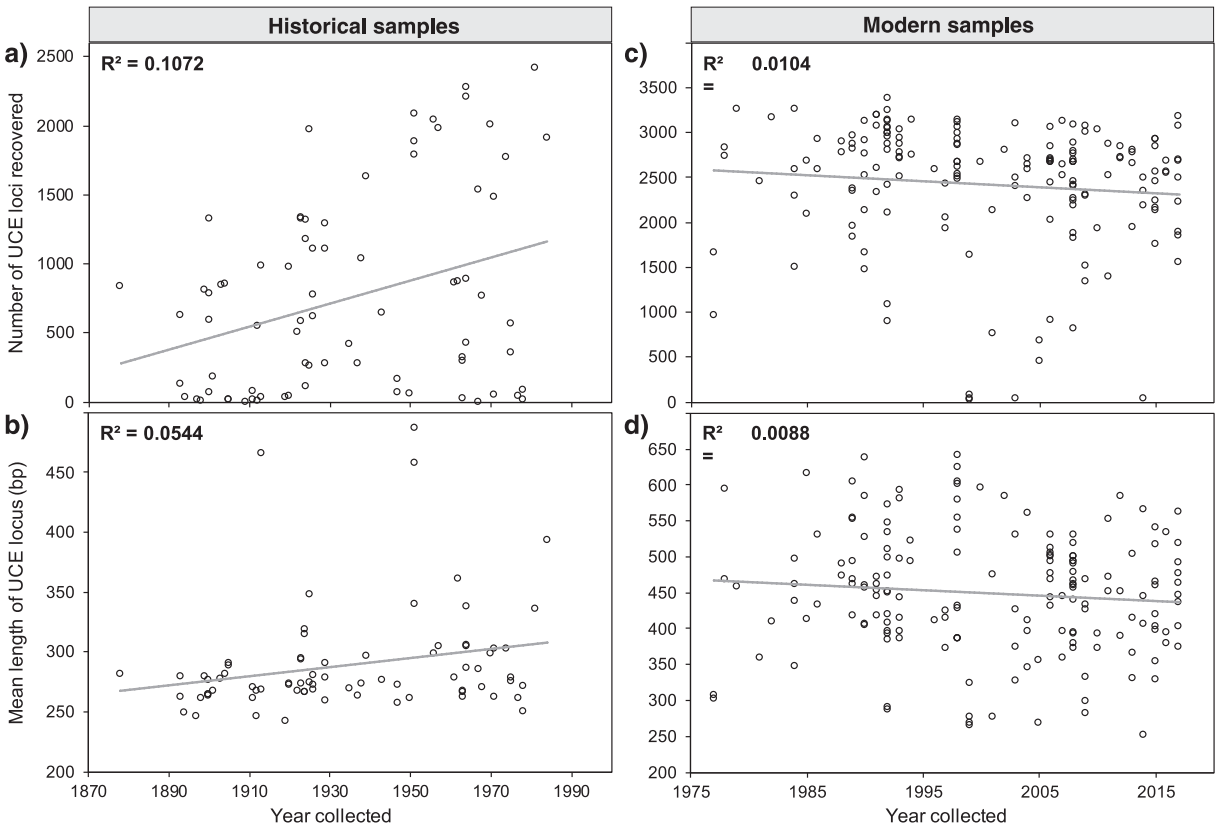

Fig. 1. Relationship between sample age and number of UCE loci recovered for historical (**A**) and modern (**C**) samples, and mean length of UCE locus recovered for historical (**B**) and modern (**D**) samples. Gray lines represent linear regressions. Analyses for historical samples were statistically significant ($p < 0.05$).

age and the number of UCE loci recovered ($R^2 = 0.0104$, $p = 0.189$; Fig. 1C) or the average UCE locus length ($R^2 = 0.0088$, $p = 0.226$; Fig. 1D). Our results showed that deeper sequencing was beneficial for recovering UCE loci for historical samples—in seven of 14 samples the number of UCE loci recovered more than doubled after the second sequencing round, and for three samples (AMNH36489, USNM303846, USNM461886) there were also substantial increases (Fig. 2A)—but this did not increase the mean length of those UCE loci (Fig. 2B).

3.2. *Topological incongruence and the influence of data filtering strategies*

To understand the impacts of dataset filtering and avoid bias that could further confound squirrel systematics, we generated six distinct datasets (see Table 1) that were analyzed with three methods of inference each, yielding a total of 18 phylogenetic analyses considering specimens as terminals in addition to 12 species trees generated with the coalescent methods (see Supplementary Fig. S1). None of the analyses considering the specimens as terminals resulted in identical topologies, as it is evident in a scatterplot of the first two principal coordinates (PC) from a PCoA and on a heat map, both based on a pairwise RF matrix, including 18 trees (Fig. 3A and B). In these among-method comparisons, IQ-Tree and SVDquartets trees are less dissimilar to each other (with a partial overlap in the first PC of the PCoA), whereas ASTRAL-III provided the most distinctive topologies according to both the two-dimensional space (Fig. 3A) and the RF scaled distances (Fig. 3B). When comparing only the species trees provided by the coalescent methods, we observed less dissimilarity among trees according to the RF values (Fig. 3D), and even complete overlapping between some trees, such as ASTRAL-III topologies resulting from 50 and 60% matrices filtered by information content, and SVDquartets topologies resulting from 40 and 50% matrices without filter for information content (Fig. 3C).

Focusing only within methods, IQ-Tree provided the most congruent topologies (mean RF distance of 0.25 ± 0.04), while ASTRAL-III recovered the most dissimilar trees (0.37 ± 0.08); the mean RF distance among SVDquartets topologies was intermediate (0.33 ± 0.08). Using IQ-Tree and SVDquartets, matrices with 40 and 50% of completeness yielded more similar trees among each other, whereas 60% matrices provided the most unique topologies (Fig. 3A and B; Supplementary Fig. S2A, B, E and F). This was true regardless of whether we analyzed unfiltered datasets or datasets containing only UCE loci with at least 5% of informative sites. On the other hand, with the gene-tree based coalescent method in ASTRAL-III, matrices filtered based on information content clearly segregated from unfiltered ones (Fig. 3A and C; Supplementary Fig. S2C). However, using a filter for information content only helped to increase topological cohesion for the species trees generated with ASTRAL-III (Fig. 3C), as the two-dimensional space occupied by topologies from datasets filtered by information content has similar size or is even bigger than the space occupied by topologies resulting from unfiltered datasets analyzed with IQ-Tree, SVDquartets, and ASTRAL-III when considering the specimens as terminals (Supplementary Fig. S2A, S2C, and S2E). The same pattern can be observed in the heat maps, where the dissimilarities among filtered datasets are equivalent or higher than the values among unfiltered ones in those methods (see Supplementary Fig. S2B, S2D, and S2F).

Regarding the nodal support, among-methods comparison shows that the median bootstrap values obtained from IQ-Tree analyses were higher than the nodal support obtained for the coalescent methods (Fig. 4). Within method comparisons demonstrate that 40 and 50% matrices recovered higher medians and shorter interquartile ranges than 60% matrices (Fig. 4). Moreover, when examining the support for all nodes (Fig. 4A, B, C), across all methods of inference datasets with no filter for information content did not vary notable from datasets with ≥ 5% of informative sites. However, when focusing on the supraspecific

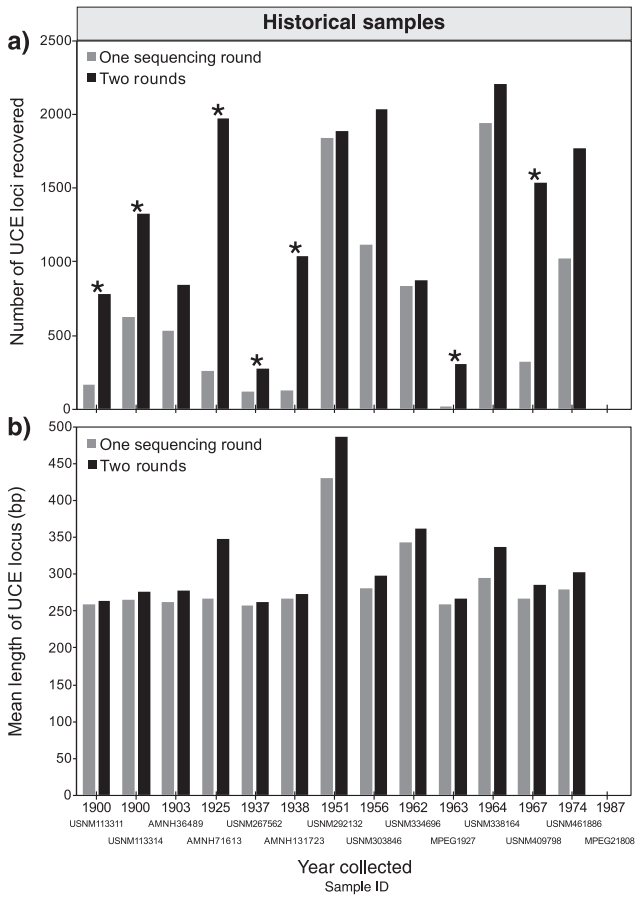

Fig. 2. Comparison of the number of UCE loci recovered for historical samples (**A**) and the mean length of UCE locus (**B**) from one sequencing round and two sequencing rounds. Samples marked with asterisk doubled the number of UCE loci after the second sequencing round.

relationships (Fig. 4D, E, F), ASTRAL-III inferences advantaged from the filter for information content, showing an increasing on the support values (Fig. 4E).

In summary, for IQ-Tree and SVDquartets, datasets including more UCE loci (40 and 50% matrices) provided the most similar topologies and the highest values of nodal support, and filtering for information content did not provide noticeable improvement in any of these metrics. For ASTRAL-III, less strict datasets (40 and 50% matrices), also yielded higher support values, but they did not provide better results on topological concordance as in the other two methods; also for ASTRAL-III, filtering out loci with <5% of informative sites helped to increase the similarity among species trees and support for interspecific relationships.

3.3. Relationships among sciurid subfamilies and remarkable structure within Sciurillinae

The great majority of our phylogenetic analyses recovered identical relationships among the five currently recognized subfamilies of Sciuridae (Fig. 5A; see also Supplementary Figs. S3–S20). The only conflicting results were between the most external branches, Ratufinae and Sciurillinae. All matrices analyzed via concatenation in IQ-Tree, except one, recovered Ratufinae as sister to all other subfamilies of Sciuridae, followed by Sciurillinae (Fig. 5A). These relationships were observed in all SVDquartets topologies, and were also the most frequently observed in ASTRAL-III (three out of six analyses). Importantly, alternative hypothesis for the initial diversification of Sciuridae from both IQ-Tree and ASTRAL-III did not receive significant support (bootstrap < 70% and local posterior probability < 0.95, respectively; Supplementary Figs. S5, S9, S10, and S12). Regarding the remaining subfamilies, all methods recovered Sciurinae sister to Xerinae + Callosciurinae with full support (bootstrap = 100%; local posterior probability = 1).

Our analyses of UCE data based on both concatenated and coalescent approaches consistently recovered two fully supported clades within the monotypic subfamily Sciurillinae (Fig. 5A). The first clade is composed of four samples from two localities in Peru, western Amazon (localities [1] and [2] in Fig. 5B); the second clade includes five samples from eastern Amazon, which are structured in regards to the Amazon river: two samples from the south bank in Pará and Amazonas, Brazil (localities [3] and [4] in Fig. 5B), and three specimens from the north bank, being two from Brazil (localities [5] and [6] in Fig. 5B), and one from French Guiana (locality [7] in Fig. 5B). Remarkably, the length of the branches subtending these two clades in IQ-Tree analyses are longer than the branches subtending most congeneric species of tree squirrels (Figs. 5A and 7), suggesting that further investigation is necessary to delimit these divergent lineages.

3.4. Phylogeny and systematics of the Sciurini radiation

Our phylogenetic inferences included 199 specimens of tree squirrels, representing 14 genera and 35 species (sensu Abreu-Jr et al., 2020b). Discrepant topologies were recovered for different datasets and inference methods also within this clade. To simplify the description of the systematic results, we consider the most inclusive dataset (3841 UCEs) that corresponds to data matrices including 40% of taxa representativeness per locus and with no filter for informative sites (trees in Fig. 6), as analyses of this dataset exhibit in general higher phylogenetic consistency and nodal support (see above "Topological incongruences and the influence of data filtering strategies"). However, we also discuss specific cases where analyses of different datasets provided distinct topologies and potentially more reliable resolution (i.e., higher nodal support) for a recalcitrant lineage.

The tribe Sciurini was recovered as monophyletic, sister to the tribe Pteromyini (flying squirrels), in all phylogenetic reconstructions with full support (Fig. 5A; Supplementary Figs. S3–S20). Concatenated and coalescent-based approaches each recovered the North American genus *Tamiasciurus* (represented by *T. douglasii* and *T. hudsonicus*) as the sister lineage to all other tree squirrel genera, followed by the Borneo endemic *Rheithrosciurus macrotis*, and the Eurasian genus *Sciurus* (represented by *S. vulgaris* and *S. lis*; Fig. 6). The subsequent cladogenetic events led to the North American genera *Hesperosciurus* (composed of *H. aberti* and *H. griseus*), and a clade composed of *Neosciurus* (represented by *N. carolinensis*) as sister to *Parasciurus* (composed of *P. niger* sister to *P. nayaritensis* + *P. arizonensis*; Fig. 6). Regarding the Neotropical taxa, which we sampled most densely, all methods suggested similar relationships for the Central American genera *Echinosciurus* (represented by *E. variegatoides*) and *Syntheosciurus* (represented by *S. granatensis*), with the former as sister to a clade composed of *Syntheosciurus* + all other lineages of the tribe (Fig. 6). Notably, the genus *Microsciurus* (apart from "*Microsciurus*", a lineage already removed from the traditional concept of *Microsciurus*; see below and Abreu-Jr et al., 2020b) was not recovered as monophyletic (Fig. 6): while *Microsciurus alfari* (from Panama) consistently appears as sister to all South American taxa, different affinities were estimated for *Microsciurus* "species 1" (from Colombia). IQ-Tree inferences recovered *Microsciurus* "species 1" as sister to the genus *Leptosciurus* (Fig. 6A), whereas most SVDquartets (Fig. 6C) and some ASTRAL-III analyses (all matrices filtered by information content) estimated this taxon as sister to all South American groups.

Among the South American genera, *Leptosciurus* was recovered as monophyletic by all IQ-Tree and SVDquartets analyses (Fig. 6A and C), and five out of six ASTRAL-III analyses. With the exception of some ASTRAL-III inferences, *Leptosciurus* was recovered as sister to *Simosciurus* + all Cis-Andean genera (Fig. 6A and C). The relationships among its

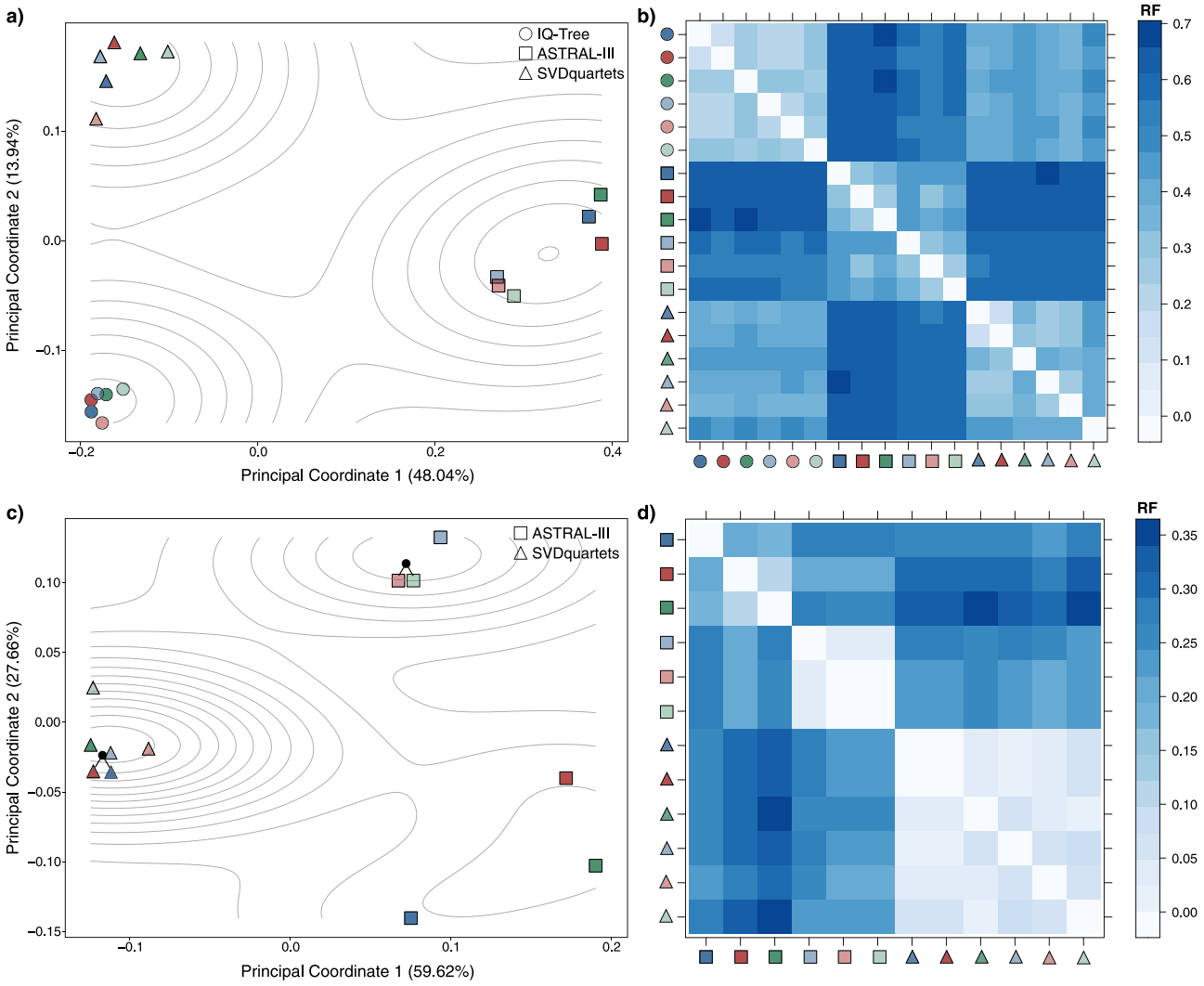

Fig. 3. Comparisons among phylogenetic trees inferred in this study. Scatterplot depicting results of Principal Coordinate Analyses (PCoA) based on matrices of Robinson–Foulds distances (RF) calculated from (**A**) trees considering all specimens as terminals and (**C**) species trees from coalescent methods. Blue symbols correspond to topologies resulting from 40% data matrices, red symbols to 50% data matrices, and green symbols to 60% data matrices. Dark shades represent datasets not filtered for information content and light shades represent dataset including only loci with ≥5% informative sites (IS). Isoclines highlight the density of trees. Heat map of normalized pairwise RF distances among (**B**) specimen trees and (**D**) species trees. Darker shades correspond to higher dissimilarity between pairs of trees. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

species, however, varied depending upon inference method (Fig. 6). The genus *Simosciurus* (represented by *S. nebouxii* and *S. stramineus*) was recovered as monophyletic and sister to a clade composed of *Guerlinguetus*, *Hadrosciurus*, and "*Microsciurus*" in all IQ-Tree analyses (Fig. 6A), and in all ASTRAL-III analyses performed with datasets filtered based on information content. The relationships among the last three South American genera were also not identical across all methods. In the concatenated approach, "*Microsciurus*" appeared as sister to a clade composed of *Hadrosciurus* and *Guerlinguetus*, while in most coalescent analyses the position of "*Microsciurus*" was switched with *Guerlinguetus* (Fig. 6). Relationships within "*Microsciurus*" and *Guerlinguetus* were identical in all methods (Fig. 6). In the genus *Hadrosciurus*, only *H. spadiceus* changed its position depending on the inference method, but both IQ-Tree and SVDquartets recovered it as diverging after *H. ignitus* (Fig. 6A and C).

Most species of tree squirrels for which we included multiple samples were recovered as monophyletic as well, with only two exceptions that warrant future systematic inquiry: (1) *Leptosciurus isthmius*, in which one individual (USNM554228 from Colombia) appeared as sister to (*L. similis* (*L. otinus* (*L. isthmus*, *L boquetensis*))) and another (USNM292133 from Panama) appeared as sister to *L. boquetensis* in the IQ-Tree topology (Fig. 7A); and (2) *Guerlinguetus aestuans* "c"—a putative unnamed species recognized by Abreu-Jr et al. (2020b)—represented here by 26 specimens that composed two distinct clades in our analyses with UCE data (Fig. 7B).

It is also notable that some species of Neotropical tree squirrels are composed of highly divergent, geographically structured lineages. For example, *Syntheosciurus granatensis* includes one clade composed of specimens from Colombia, Ecuador, Peru, Venezuela, and Trinidad and Tobago, and another clustering specimens from Panama (Fig. 7A). "*Microsciurus*" *flaviventer* includes at least four parapatric lineages along the Amazon basin (Fig. 7A; Supplementary Table S1). *Hadrosciurus ignitus* exhibit one lineage including specimens from Brazil and Peru, and another composed of specimens from Argentina and Bolivia (Fig. 7A). *Guerlinguetus brasiliensis* also shows a deep geographic structure, with a clade assembling specimens from southeastern Brazilian Amazon (samples from Pará) and another clustering specimens from the eastern portion of Brazil, along the Atlantic Forest (Fig. 7B).

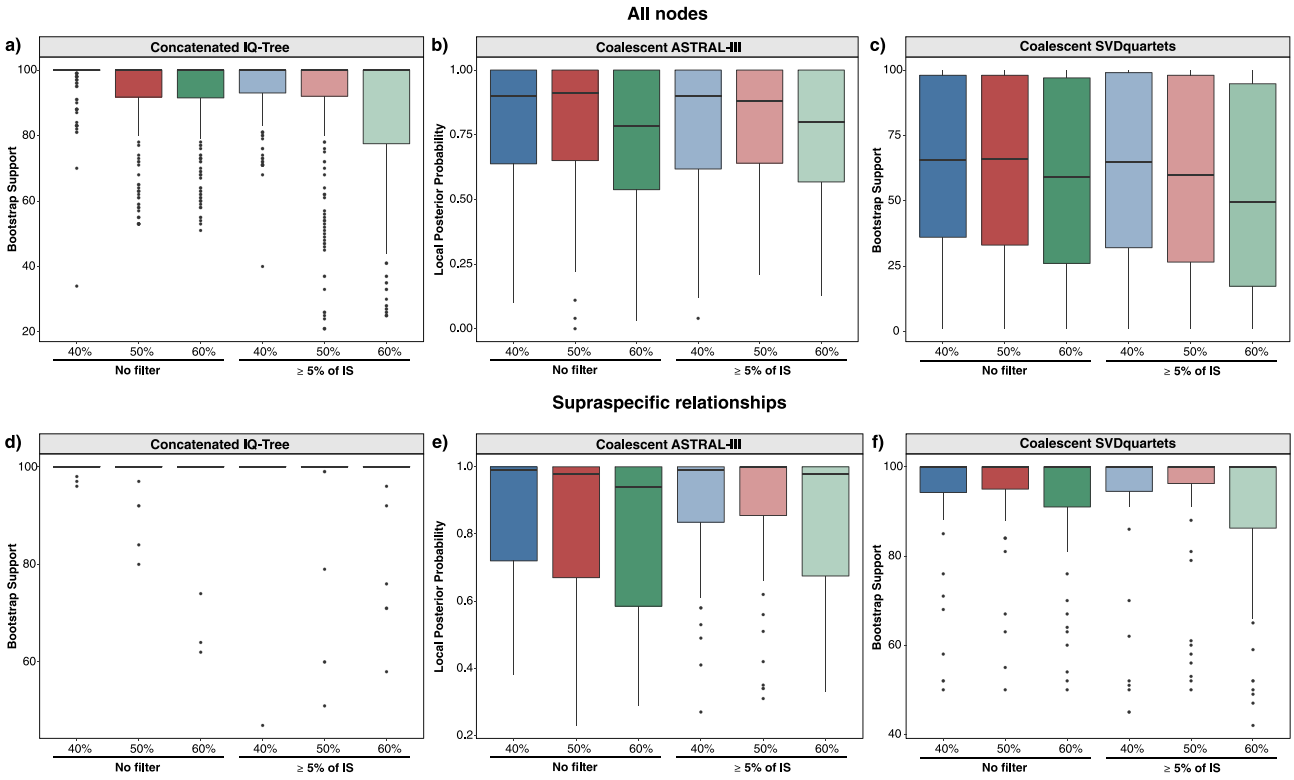

Fig. 4. Boxplots showing changes in overall support values (**A–C**) and in the nodal support for interspecific relationships (**D–F**) of trees inferred for each dataset using three distinct methods: IQ-Tree, ASTRAL-III, and SVDquartets.

4. Discussion

4.1. Using historical samples to unpuzzle the tree of life

Squirrels are a globally-distributed clade that has diversified to inhabit a wide variety of environments and ecological contexts, despite being relatively conserved in morphological characters (McLean et al., 2018; Zelditch et al., 2015). The sciurine mode of radiation, however, has been taken to an extreme in the Neotropics, where genomic data demonstrate tremendously rapid lineage accumulation, and that taxonomic understanding is still in flux at multiple levels. In a previous contribution, Abreu-Jr et al. (2020b) showed the pivotal role of historical samples in the phylogenetic inferences of the tribe Sciurini, based on mitochondrial genome data. About a third of the samples used in that study were historical; likewise, about 20% of the samples in our UCE-based phylogenies are historical, and this provided critical material for 11 nominal taxa of tree squirrels (almost a third of Sciurini species analyzed in this study) for which there were no modern samples available (see Figs. 6 and 7). Also, for five additional species, at least half of the samples we used came from historical specimens and were key for understanding species limits (see Fig. 7). Finally, the ability to recover UCEs from old museum specimens allowed us to include an important historical specimen of the enigmatic and monotypic *Sciurillus pusillus* from [River Arataye, Inini] French Guiana, which represents our closest sample to the type locality of this species ("Cayenne", French Guiana; Vivo and Carmignotto, 2015).

While sample age did not impact recovery of mitochondrial genome data for historical samples (Abreu-Jr et al., 2020b), we found major effects of age on enrichment of UCE data for those same samples (Fig. 1). This corroborates previous studies indicating that the number and average length of UCE loci recovered decreases with specimen age (Heintzman et al., 2014; McCormack et al., 2016; McDonough et al., 2018). However, despite the many studies that have utilized historical samples for UCE-based phylogenomics, the extent to which sample age and preservation method impact UCE recovery, as well as downstream inferences, has been unclear for most vertebrate groups. Some insights arose from studies of invertebrate taxa (Blaimer et al., 2016; Branstetter et al., 2021; Derkarabetian et al., 2019; Ruane and Austin, 2017), but those organisms are submitted to different preparation procedures and preservation methods compared to mammals. The decrease in the mean length of UCE loci we observed is likely a result of extensive DNA fragmentation, which is accelerated with age if the DNA is not properly stored (McDonough et al., 2018). McCormack et al. (2016) suggested that extra sequencing could help to improve the number of UCE loci recovered for historical samples but not average locus length. Our results with empirical data clearly corroborate these suggestions (see Fig. 2), and we demonstrate that locus length is a critical parameter for coalescent-based inference (mainly gene-tree based methods; see more details in the next section).

Recently, a suggestion that natural history museums should cut off a fragment of unique specimens—representing taxa with no fresh tissue available—for cryogenic preservation has been raised in the literature (McCormack et al., 2016). Given that natural history museums are critical repositories of global biodiversity, there is a need to weigh the feasibility of such an effort against the benefit for biodiversity science. Although we are aware of the potential damage to voucher material, our results suggest that this procedure would help decelerate DNA degradation—particularly for specimens collected in the recent past and that were not submitted to severe chemical treatments (e.g., arsenic)—and, by extension, to increase the likelihood of obtaining good-quality nuclear DNA for irreplaceable historical samples, including type specimens, that would be critical for appropriate taxon name attribution in the genomic era. This may be necessary because, even for a widespread and charismatic taxon such as tree squirrels, substantial taxonomic and geographic sampling gaps coupled with more restrictive permits to collect specimens in nature are constraining. This will continue to be the case in the future, limiting our ability to fully resolve the Tree of Life.

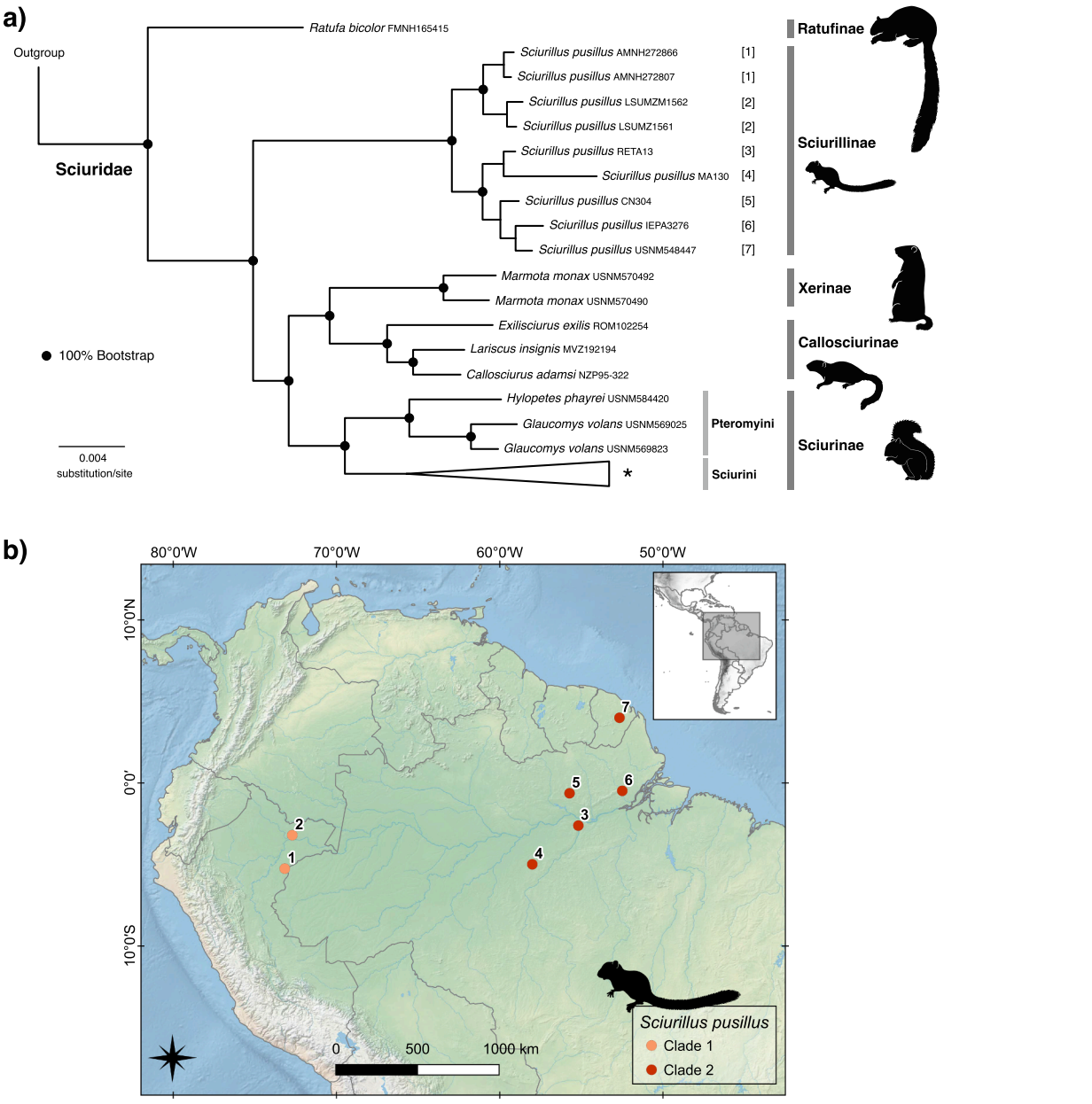

Fig. 5. (**A**) Phylogenetic hypothesis for the inter-subfamily relationships of Sciuridae based on a Maximum Likelihood analysis (IQ-Tree) performed with Dataset 1. Samples are identified by the binomial, followed by the voucher catalog number. For Sciurillinae, the numbers at the tips of the terminals correspond to localities shown on B. *Terminals of the tribe Sciurini were collapsed. (**B**) Map of collecting localities of samples of *Sciurillus pusillus* in the Amazon basin.

4.2. Impact of data filtering and inference method on resolving rapid radiations

Optimizing filtering approaches for genome-wide datasets is critical in modern phylogenomics, as these datasets routinely contain missing taxa, missing loci, or both, which can impact inferences where incomplete lineage sorting, early lineage isolation or introgressive hybridization are present (Blom et al., 2017; Hosner et al., 2016; Platt et al., 2018). There is no consensus in the recent literature on what filtering strategy is optimal for a particular method of inference, as the clade of study and biological properties of the loci being analyzed play a major role. Therefore, when estimating phylogeny of recalcitrant lineages, it is essential that researchers examine the influence of a variety of filtering parameters across conceptually distinct inference methods to identify zones of methodological inconsistency (McLean et al., 2019; Platt et al., 2018).

We created matrices that varied in sample representativeness per loci (e.g., 40, 50 and 60%), expecting that the inclusion of a greater number of loci, even with the consequent reduction on the number of samples per locus, would positively affect the robustness of our inferences, based on previous studies (e.g., Hosner et al., 2016; Streicher et al., 2016; Platt et al., 2018). Indeed, our results confirmed that datasets containing more loci (40% and 50% matrices) performed better than the stricter dataset (60% matrix) in terms of nodal support values and their variation (Fig. 4). Increasing the number of loci in UCE datasets has been shown to be effective at maximizing the statistical support and also at improving topological consistency (McLean et al., 2019; Platt et al., 2018). Our results on smaller RF distances among topologies obtained from more inclusive datasets, especially when estimated with IQ-Tree and SVDquartets, also corroborated this last assumption (see Fig. 3B).

Regarding the second filtering strategy (information content), the initial matrices were curated to include only UCE loci with a minimum

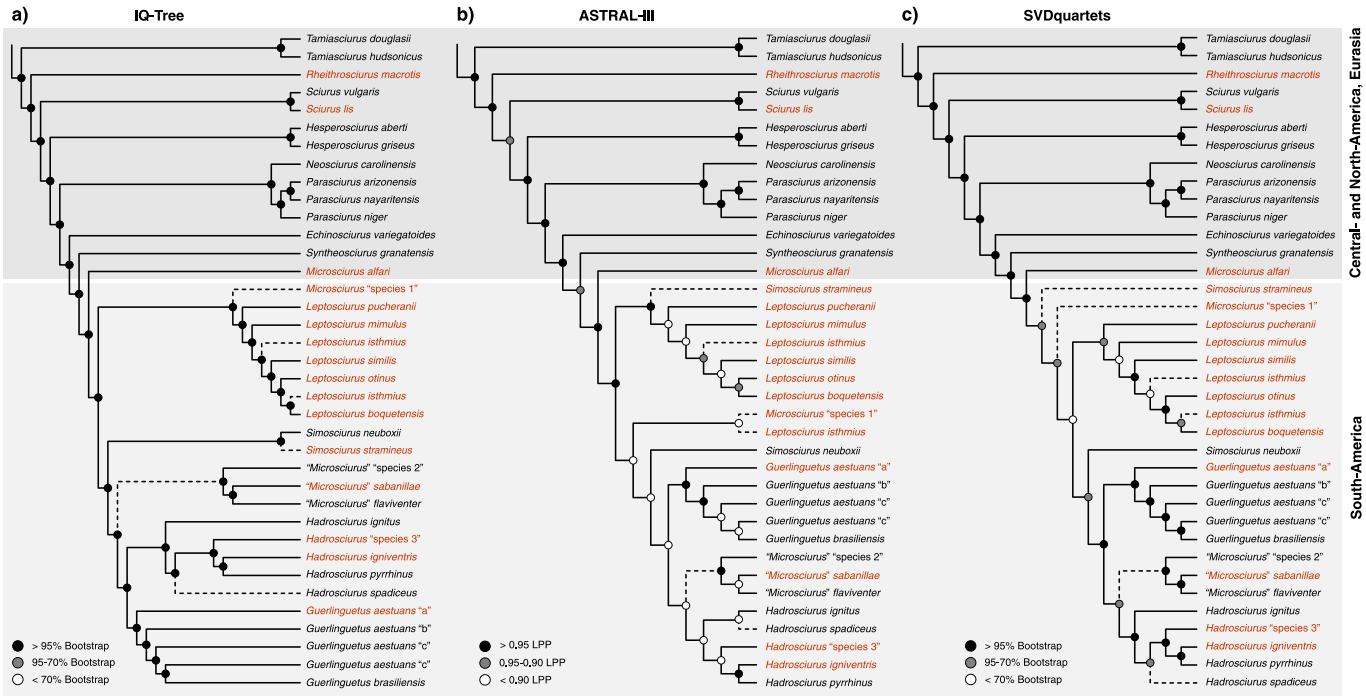

Fig. 6. Phylogenetic hypotheses for the species-level relationships of the tribe Sciurini recovered by IQ-Tree, ASTRAL-III, and SVDquartets. Analyses were performed with Dataset 1. Taxonomic identifications at both genus- and species-level follow Abreu-Jr et al. (2020b). Dashed branches correspond to conflicting relationships among methods. Species names colored with red represent lineages composed exclusively or dominated (50% or more of the specimens) by historical samples. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

of 5% of informative sites, and two major findings emerged from this exercise. First, this approach did not impact inference on the largest datasets (40 and 50% matrices), but did negatively impact the 60% matrix when analyzed with IQ-Tree and SVDquartets, as higher dissimilar topologies with lower overall nodal support were recovered within and among methods upon this filter (Figs. 3 and 4). This is in contrast to other studies in which filtering loci by information content has helped to improved topological concordance among these methods (Hosner et al., 2016; Manthey et al., 2016) even if this improvement did not result in full concordance (McLean et al., 2019). We therefore suggest that the disadvantage of filtering so many loci (with <5% of informative sites) was perhaps greater than the advantage of analyzing fewer (but supposedly more phylogenetic informative) ones.

Second, this strategy positively impacted ASTRAL-III inferences on all datasets, as higher nodal support values were obtained, particularly for intraspecific relationships (Fig. 4). Also, ASTRAL-III topologies inferred from datasets filtered by information content were less dissimilar to those estimated by IQ-Tree and SVDquartets (Fig. 3). Since coalescent methods are sensitive to inclusion of imprecise gene trees, as those resulting from very conserved locus with low substitution rates (Blom et al., 2017; Hahn and Nakhleh, 2016; Mirarab et al., 2016), we confirmed that ASTRAL-III analyses benefited from datasets with higher rates of informative sites. It is important to mention that our filtering approach based on information content did not take into account the compositional heterogeneity or the clock-like behavior of the datasets, and particularly the first has the potential to influence the resolution of difficult nodes increasing phylogenetic discordance when estimated from large multi-locus datasets (Sharma et al., 2015). Compositional heterogeneity is expected in datasets with non-uniform substitution patterns resulted from nonstationary evolution (Sheffield, 2013). The sciurine mode of radiation with impressively rapid lineage accumulation in some areas of the globe, as in the Neotropics (Abreu-Jr et al., 2020a; Zelditch et al., 2015), might be a sign that this phenomenon could be a source of noise on the establishment of phylogenetic relationships in the group, thus future studies on squirrels phylogenomics should test for this putative bias.

Comparing the optimality criteria employed, IQ-Tree and SVDquartets inferred slightly more similar topologies, while ASTRAL-III suggested the most divergent ones (Fig. 3). ASTRAL-III, as a two-step coalescent method that uses a collection of gene trees to infer a species tree (Zhang et al., 2018), may have poorer performance given the limited length of some UCE loci recovered (the mean length of our UCE loci was 437 bp; range 267–640) and thus poorly resolved gene trees. According to Chou et al. (2015), gene trees estimated on short alignments are prone to high estimation error, and this could be more pronounced in UCEs since core regions are targeted but are much less variable than the flanking regions (Faircloth et al., 2012; Platt et al., 2018). Previous studies, analyzing UCEs with similar mean length or even loci with longer average length, also recovered inconsistent and heterogeneous results from ASTRAL inferences (e.g., Bryson et al., 2016; Baca et al., 2017; Streicher and Wiens, 2017), corroborating that UCE loci are relatively short markers—in general with small percentages of informative characters—making it difficult to produce well-resolved gene trees (Platt et al., 2018). Additionally, we employed RAxML to generate the individual gene trees, and it has been shown that this method might infer arbitrarily resolved clades—especially when there are few informative characters in relation to the number of terminals—which can produce extraneous conflicting gene trees, negatively impacting the inference of the species-tree (Simmons et al., 2016; Simmons and Gatesy, 2021).

SVDquartets is consistent with the multispecies coalescent but differs from ASTRAL-III in that it takes unlinked site data to infer quartet trees for subsets of four terminals, and then combines the quartet trees onto a species tree (Chifman and Kubatko, 2014). Since this site-based summary method does not require representing each gene by a single tree, it is not sensitive to individual gene tree estimation error (Chou et al., 2015). This may help to explain why SVDquartets resulted in topologies largely diverging from ASTRAL-III. Disagreement between coalescent methods is frequently reported and for inferences of recalcitrant clades it can be persistent (Chou et al., 2015; Linkem et al., 2016; McLean et al.,

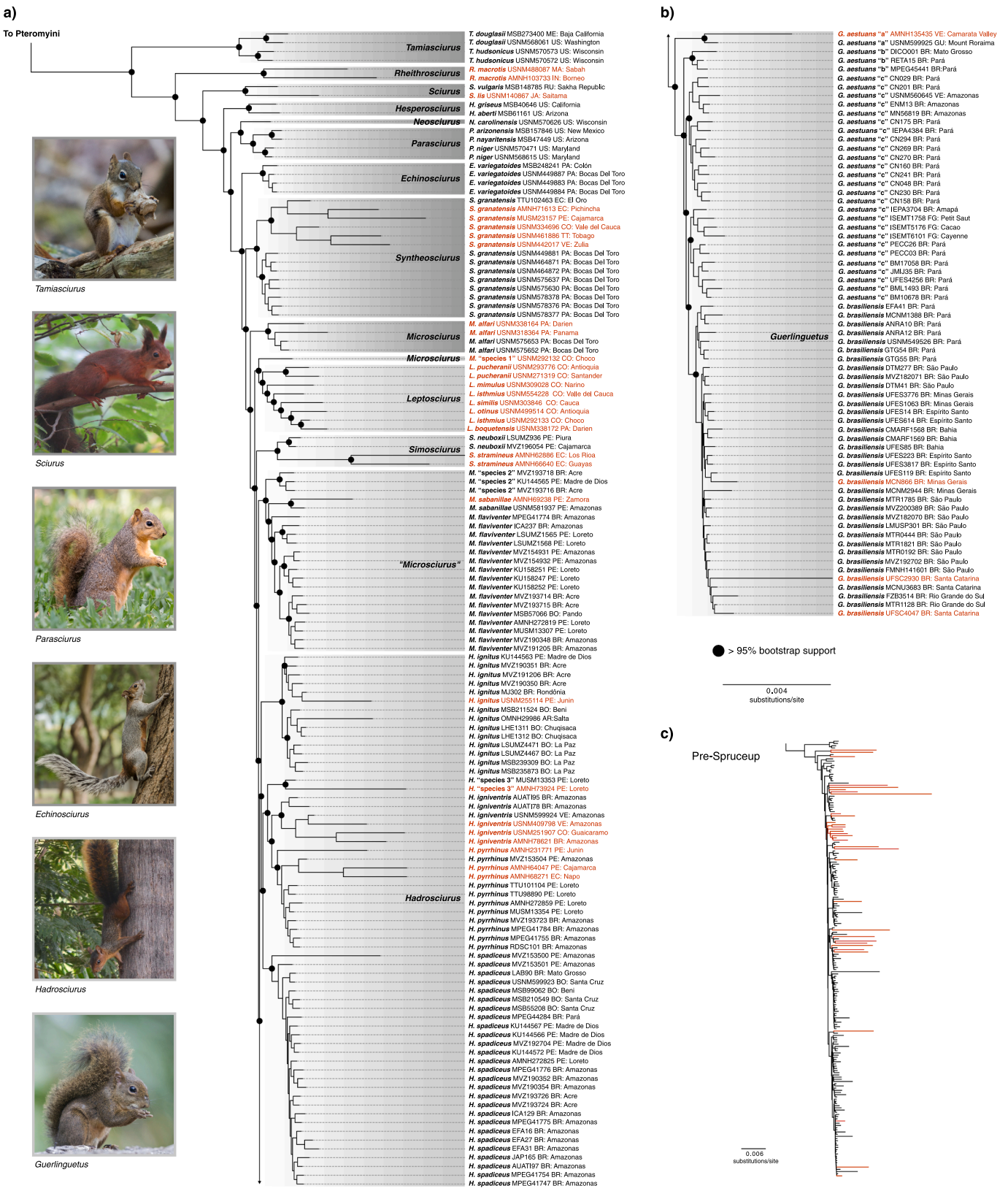

Fig. 7. (**A and B**) Detailed phylogenetic hypothesis of the tribe Sciurini based on a Maximum Likelihood analysis (IQ-Tree) performed with a concatenated matrix of 3841 UCE loci (40% taxa representativeness per locus and no filter for information content). Terminals are named with binomials (following Abreu-Jr et al., 2020b), accompanied by museum voucher numbers and geographic information (country code and state/department). Historical samples are colored with red. (**C**) Simplified phylogenetic tree estimated with IQ-Tree on the same dataset but prior to processing in SPRUCEUP, highlighting terminal represented by historical samples. Photography credits: S. Oliveira (*Tamiasciurus*), H. Garbino (*Sciurus*), P. Peloso (*Parasciurus* and *Echinosciurus*), F. Abra (*Hadrosciurus*), and F. Arantes (*Guerlinguetus*). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

2019; Streicher et al., 2016).

The last inference method used here, IQ-Tree, uses a concatenated alignment of all loci to estimate a species tree based on the ML algorithm (Minh et al., 2020). This method provided the least discordant set of topologies with the highest medians of nodal support. The inflation of support values has been observed in a variety of concatenation-based studies, and in our case is probably a response of prioritizing overall character sampling over maximizing taxon-sampling per locus in the largest (40 and 50%) data matrices (Platt et al., 2018; Streicher et al., 2016).

4.3. A phylogenomic hypothesis for subfamilies of Sciuridae

This is the first phylogenomic study including all five currently recognized Sciuridae subfamilies (sensu Thorington et al., 2012). Previous studies, based on a few mitochondrial and nuclear genes (e.g., Mercer and Roth, 2003; Steppan et al., 2004) or on supermatrices with up to 31 genes (e.g., Fabre et al., 2012; Zelditch et al., 2015; Upham et al., 2019), have estimated conflicting relationships for those subfamilies, especially among lineages that diverged early such as the Neotropical pigmy squirrel (Sciurillinae) and the Asian giant tree squirrels (Ratufinae). Mercer and Roth (2003) recovered Sciurillinae as sister to the rest of Sciuridae; in Fabre et al. (2012), Sciurillinae diverged after Ratufinae; and in some other inferences Sciurillinae and Ratufinae appear as sister subfamilies (Steppan et al., 2004; Zelditch et al., 2015; Upham et al., 2019).

All of our most strongly-supported inferences place the Asian giant squirrels as the sister group of all extant squirrels, followed by Sciurillinae (see Fig. 5A), as suggested by Fabre et al. (2012). Alternative hypotheses for the initial diversification of Sciuridae (recovered mostly by ASTRAL-III) were always poorly supported. Relationships among the remaining subfamilies were identical among methods and with strong nodal support in the 18 analyses performed in this study. UCE datasets suggest Sciurinae (including Sciurini and Pteromyini) as the sister group of a clade composed of Callosciurinae (southern Asian tree squirrels) + Xerinae (African and Holarctic ground squirrels and African tree squirrels) (Fig. 5A). These phylogenetic affinities are similar to previous studies (Fabre et al., 2012; Steppan et al., 2004; Upham et al., 2019) and provide a sound basis for future comparative work.

4.4. Cryptic diversity within Sciurillinae

Based on DNA sequences from one nuclear and two mitochondrial genes, Mercer and Roth (2003) have estimated a late Miocene divergence between two individuals of *Sciurillus pusillus* from opposite ends of the species geographic range, firstly implying a great genetic diversity within this enigmatic group of squirrels. Our analyses strikingly support and escalate this interpretation, as we consistently recovered at least two highly divergent lineages within *Sciurillus*. Although our sampled specimens are also from opposite extremes of the Amazon basin (see map on Fig. 5B) and this could be interpreted as a sampling bias, the current known geographical range of this taxon actually shows a disjunct distribution, with a large hiatus in the central Amazon (Calderón-Capote et al., 2015; Vivo and Carmignotto, 2015). Vivo and Carmignotto (2015) examined the morphological variation, mostly external and craniodental features, of *S. pusillus* and find no clear geographically structured variation among populations. By contrast, Calderón-Capote et al. (2015) documented conspicuous differences in the baculum morphology between specimens from the western Amazon (Colombia) and eastern Amazon (Brazil), suggesting that this taxon might not be monotypic, as also evidenced by our genomic data.

The indication of a greater diversity within this ancient lineage provides new incentive to unravel the diversity, biogeographic history and evolution of sciurids worldwide. Currently, the origin of all squirrels has been estimated as Eurasia (Rocha et al., 2016), which is home to four extant species of *Ratufa*—another deep branch of Sciuridae (Burgin et al., 2018; Upham et al., 2019). Our results showing that the diversity of ancient Neotropical lineages is likely underestimated might question previous inferences on the biogeographic origin of this group, and highlight the crucial role of the Neotropics on the initial diversification of sciurids.

4.5. Discordances on Sciurini phylogenomics and the influence of sample age

Molecular systematics of the tribe Sciurini have been subject of many phylogenetic studies, but most of those emphasized Eurasian (Aghbolaghi et al., 2020, 2019; Oshida et al., 2009; Oshida and Masuda, 2000), North American (Hope et al., 2016), or Mesoamerican (Villalobos and Cervantes-Reza, 2007; Villalobos and Gutierrez-Espeleta, 2014) species, and others included a broader geographic coverage of tree squirrels (Fabre et al., 2012; Pečnerová et al., 2015; Pečnerová and Martínková, 2012; Zelditch et al., 2015). However, the majority of tree squirrel diversity lies in the Neotropical radiation, and only one recent study comprehensively sampled the South American forms (Abreu-Jr et al., 2020b). Based on mitogenome data, Abreu-Jr et al. (2020b) estimated that 46 species are included in Sciurini, and proposed the use of 14 generic names, employed by previous authors (e.g., Allen 1915; Moore 1959; Vivo and Carmignotto, 2015), to represent the genus-level diversity of the group.

Our analyses of UCE data include 35 of these putative species (sensu Abreu-Jr et al., 2020b) representing all 14 genera, a broader taxonomic representation than previous studies including nuclear markers (e.g., Fabre et al., 2012; Pečnerová et al., 2015; Zelditch et al., 2015). We observed two distinct patterns across inference methods. For the older and less diverse (monotypic or with few species) lineages—including North American (*Tamiasciurus*, *Hesperosciurus*, *Neosciurus*, and *Parasciurus*) and Eurasian (*Rheithrosciurus* and *Sciurus*) genera—that have been estimated to diversify between 14 and 7 Mya (Abreu-Jr et al., 2020a), and also for two Central American genera (*Echinosciurus* and *Syntheosciurus*), all methods recovered identical relationships with full support (Fig. 6). This portion of the tree squirrel phylogeny is not only consistent among UCE inferences, but also mostly corroborates the mitogenomic hypothesis proposed by Abreu-Jr et al. (2020b)—except for *Echinosciurus* and *Syntheosciurus*, for which mitogenomic data recovered a sister relationship (see below). Moreover, most of the samples representatives of these taxa are modern and yielded high numbers of UCE loci with relatively longer lengths (see Fig. 6 and Supplementary Table S2).

Regarding the remaining Central and South American taxa—which experienced a more recent and extremely rapid diversification after the South American invasion around 6 Mya (Abreu-Jr et al., 2020a)—substantial conflict remains at both genus- and species-levels. *Microsciurus* (sensu Abreu-Jr et al., 2020b) was consistently not recovered as monophyletic, and *Microsciurus* "species 1" was recovered with distinct phylogenetic affinities depending upon inference method (see Fig. 6). In the mitogenome analyses, *Microsciurus* was found as sister to *Echinosciurus* + *Syntheosciurus*, with the three genera composing a large predominantly Central American clade (Abreu-Jr et al., 2020b). This is perhaps the most conflicting point between the UCE and the mitogenome phylogenies, and also the most poorly represented in term of taxonomic coverage in the UCE analyses, as one species of *Syntheosciurus* and four species of *Echinosciurus* were missing in the latter approach.

The Andean/Trans-Andean genus *Leptosciurus*, represented in our analyses exclusively by historical samples (Fig. 6), exhibited conflicting interspecific relationships among inference method, which also differ from the mitogenomic hypothesis (Abreu-Jr et al., 2020b). Besides detecting this conflicting region in a portion of the Sciurini tree dominated by old samples, we observed that the mean nodal support of branches subtending historical specimens across all methods are lower (although not statistically significative; $p > 0.05$ according to Mann-Whitney U tests) than the mean values obtained for branches

subtending modern samples (Fig. 8). Within the Trans-Andean genus *Simosciurus*, *S. stramineus* (represented exclusively by historical individuals) varied its position in ASTRAL-III inferences performed with unfiltered datasets and also in all SVDquartets analyses. However, matrices filtered based on information content analyzed with ASTRAL-III and IQ-Tree recovered *S. stramineus* and *S. neuboxii* as sister taxa, and sister to the Cis-Andean genera, identical to the mitogenome inferences (Abreu-Jr et al., 2020b). Thus, we believe that issues related with sample age and, by extension, lack of phylogenetically-informative data (i.e., from UCEs flanking regions), prevented us to obtain more consistent resolution for the position of *S. stramineus* in the coalescent approaches.

The relationships among the South American Cis-Andean genera *Guerlinguetus*, *Hadrosciurus*, and "*Microsciurus*" were identically inferred by the coalescent methods, which are in full agreement with the mitogenomic hypothesis (Abreu-Jr et al., 2020b). Within *Guerlinguetus* and "*Microsciurus*" no disagreement was found among UCE inferences nor between nuclear and mitochondrial data. Within *Hadrosciurus*, ASTRAL-III was the only method to suggest a sister relationship between *H. spadiceus* and *H. ignitus*, as was recovered in the mitogenome inferences (Abreu-Jr et al., 2020b). The consistent phylogenetic results obtained for these widely distributed Cis-Andean taxa may reflect our dense sampling of modern specimens including important geographic variants (see Fig. 7A).

Our UCE-based phylogenetic inferences corroborated most species-level lineages recognized by the mitogenome data. The only cases of non-monophyly among all species of tree squirrels included in this study are *Leptosciurus isthmius* and *Guerlinguetus aestuans* "c" (Fig. 7). As a matter of fact, the two *L. isthmius* specimens represented in our UCE datasets are the most divergent in the mitogenome phylogeny (Abreu-Jr et al., 2020b). Further analyses including larger sample sizes are needed to evaluate the species limits within this taxon. *Guerlinguetus aestuans* "c" is part of a species complex associated to the polytypic *G. aestuans* (sensu Vivo and Carmignotto, 2015). This putative species, according to Abreu-Jr et al. (2020b), includes specimens from the Atlantic Forest in north-eastern Brazil, and from a broad area in the Amazon basin (Supplementary Fig. S21A). However, those specimens were recovered as two well-supported, non-sister clades in our analyses of UCE data (Fig. 7 and Supplementary Fig. S21B). Although conflicting species recognition could simply reflect error in phylogenetic estimation (Huang et al., 2010), it also may be driven by genuine discordance between nuclear and mitochondrial marker types that, by extension, could be associated with several biological factors, such as gene flow and introgression (Platt et al., 2018; Chan et al., 2020). D'Elía et al. (2019) reviewed how hybridization and introgression have been addressed in rodent systematics, highlighting our poor understanding of the presence and magnitude of these phenomena in Neotropical taxa. High levels of introgression have been found in some lineages of mammals from South America, for example in felids (Trigo et al., 2013) and canids (Tchaicka et al., 2016), so further investigation is sorely needed to accurately delimit potentially-hybridizing lineages in groups such as *Guerlinguetus*.

Mito-nuclear discordances on the Sciurini phylogeny may also attest to additional phenomena acting on tree squirrel evolution, such as incomplete lineage sorting. The Neotropical radiation of tree squirrels has been explosive over the past 6 Mya (Abreu-Jr et al., 2020a), and this increases the probability that incomplete lineage sorting is driving at least some of the discordance we observed, including the variable topologies upon dataset filtering and optimality criterion. Conversely, the lower speciation rates along the initial diversification of Sciurini (Abreu-Jr et al., 2020a), would make it less susceptible to the effects of incomplete lineage sorting (see Cai et al., 2021; Platt et al., 2018), and it might reflect the robustness of our inferences for the deeper branches of the phylogenetic tree of tree squirrels.

5. Conclusions

As the global climate crisis increasingly places animal and plant species at extinction risk, our ability to reconstruct the Tree of Life for the tropics and infer the processes that have generated it relies (and will continue to do so in the future) heavily on historical samples curated in natural history collections. In our study, both the recovery of UCE loci and the mean length per locus were negatively impacted by sample age; we also empirically demonstrated that additional sequencing fails to mitigate decreases in the mean length of loci recovered for historical samples. Nevertheless, we resolved critical regions of the squirrel phylogeny and found that filtering based on sample presence (matrices with 40 and 50% of taxa representativeness per loci) was more important than filtering for information content for concatenation and for the coalescent site-based method. Conversely, filtering based on information content (including only loci with a minimum of 5% of informative sites) provided better topological resolution and increased nodal support for inferences with the coalescent gene-tree based method, likely due to more informative gene trees. We resolved the deeper relationships in Sciuridae (including among five currently recognized subfamilies), and also the relationships among the deepest branches of Sciurini, but conflicting relationships at both genus- and species-level remain for the rapid Neotropical tree squirrel radiation. Our results endorse that even with a tremendous amount of genome-scale sequence data, consensus can be difficult and heavily influenced by the age of available samples and the bioinformatic workflows used to optimize dataset properties. Finally, our results advance the current phylogenetic and systematic knowledge of a largely neglected group of mammals, corroborating the most recent supra-specific classification of tree squirrels and unlocking unexpected species-level diversity within both Neotropical subfamilies, which underscores the urgent need for taxonomic studies describing the Neotropical mammal diversity.

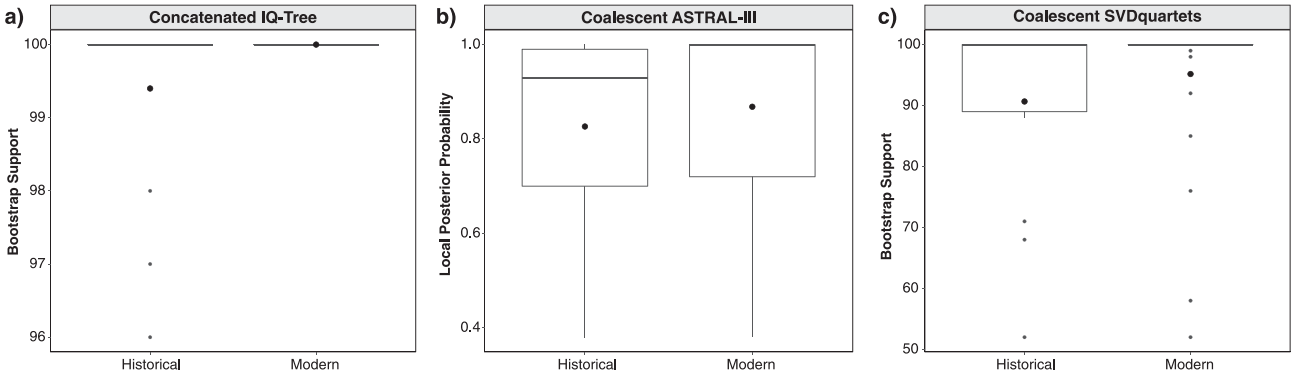

Fig. 8. Boxplots comparing nodal support for interspecific relationships between historical and modern samples from trees inferred for Dataset 1 (see Table 1) using three distinct methods: IQ-Tree, ASTRAL-III, and SVDquartets.

CRediT authorship contribution statement

Edson F. Abreu: Conceptualization, Methodology, Formal analysis, Investigation, Data curation, Writing – original draft, Visualization, Project administration, Funding acquisition. **Silvia E. Pavan:** Conceptualization, Methodology, Investigation, Data curation, Writing – review & editing, Project administration, Funding acquisition. **Mirian T.N. Tsuchiya:** Methodology, Data curation, Writing – review & editing. **Bryan S. McLean:** Conceptualization, Methodology, Data curation, Writing – review & editing. **Don E. Wilson:** Writing – review & editing, Resources, Funding acquisition. **Alexandre R. Percequillo:** Conceptualization, Writing – review & editing, Resources, Funding acquisition. **Jesús E. Maldonado:** Conceptualization, Writing – review & editing, Resources, Funding acquisition.

Declaration of Competing Interest

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

Data availability

All datasets and phylogenetic trees generated in this study are available in the figshare repository (https://doi.org/10.6084/m9.figshare.20038628.v1 and https://doi.org/10.6084/m9.figshare.20039267.v1, respectively). Illumina reads sequenced for each sample are available on GenBank under the BioProject PRJNA847638.

Acknowledgments

We would like to express our gratitude to the following curators and collection support staff, who loaned us preserved tissue samples, allowed us to analyze and obtain destructive samples from valuable historical specimens under their care, or provided essential information regarding vouchers: R.S. Voss, E. Hoeger, and N. Duncan (AMNH); I. Moya (CBF); M.R. Alvarez (CMARF); B.D. Patterson and A.W. Ferguson (FMNH); J. Valsecchi and I. Junqueira (IDSM); C.R. Silva and A.F. Sobrinho (IEPA); F. Catzeflis (ISEM); R.M. Timm and M. Eifler (KU); D.L. Dittmann, R.T. Brumfield, and J. Esselstyn (LSUMZ); E. Pasa (MCN-FZB); C.G. Costa (MCN-M); A.U. Christoff, F.B. Peters, and A.M. Gandini (MCNU); J.A. Oliveira and M. Weksler (MN); J. Silva Junior and A.M.R. Bezerra (MPEG); M. Campbell, J. Cook, and J. Dunnum (Museum of Southwestern Biology, University of New Mexico); M.T. Rodrigues, S. Baroni, and B.M.S. Batista (MTR); V. Pacheco (MUSM); C. Conroy, J.L. Patton, and F.J.M. Pascal (Museum of Vertebrate Zoology, University of Calirnia, Berkeley); M. Vivo, L.F. Silveira, and J.G. Barros (MZUSP); J.K. Braun and B.S. Coyner (OMNH); H. Garner, R.D. Bradley, and C. Phillips (TTU); L.P. Costa, Y.L.R. Leite, and M.P. Nascimento (UFES-CTA); R. Rossi and T. Semedo (UFMT); A.C. Mendes Oliveira (UFPA); J. Cherem, M. Graipel, and E.C. Grisard (UFSC); D.P. Lunde, S.C. Peurach, N.R. Edmison, I. Rochon, M. Krol, J. Jacobs, C. Ludwig, C. Huddleston, D. DiMichele, M. Braun, L.H. Emmons, and A.L. Gardner (USNM). In addition, we would like to acknowledge the Ambrose Monell Cryo Collection at the AMNH, NY, for their support in our research. A. Ravetta, G.T. Garbino and L.P. Godoy provided tissue samples or allowed us to sample recently collected or uncatalogued specimens. We thank L. Parker, M. Venkatraman, S. Castañeda and N. McInerney, who provided crucial help during lab work. We are also thankful to J.R. Prado, who shared scripts and insights on computational procedures. We are indebted to M. Hawkins, who provided UCE data for Callosciurinae. We are grateful to R. Dikow, who was always supportive with any matter regarding the Smithsonian HPC. We thank G. Libardi, who generously made the squirrel drawings depicted on Fig. 5, and S. Oliveira, H. Garbino, P. Peloso, F. Abra, and F. Arantes, who allowed us to use the photographs depicted on Fig. 7. We finally would like to thank M.P. Simmons and one anonymous reviewer for suggestions that improved the quality of this manuscript.

Funding

This work had financial support from the Conselho Nacional de Desenvolvimento Científico e Tecnológico through doctoral fellowships to EFA (147145/2016-3, 203692/2017-9, and 165553/2017-0), through a postdoctoral fellowship to SEP (302204/2020-2), and through a productivity scholarship to ARP (304156/2019-1); from the American Society of Mammalogists through the Latin American Student Field Research Award to EFA, and through the Oliver P. Pearson Award to SEP; from the American Museum of Natural History through a Collection Study Grant to EFA; from the Smithsonian Institution through postdoctoral fellowships to SEP and MTNT (Smithsonian Women's Committee), and through research funds provided to SEP, MTNT, DEW and JEM; from the National Geographic Society through an Explorers Grant to SEP (NGS-381R-18) and through Support for Women + Dependent Care to SEP (NGS 3758); and from the Fundação de Amparo à Pesquisa do Estado de São Paulo through research funds provided to ARP (09/16009-1). EFA is currently supported by the Gerstner Postdoctoral Fellowship from the Richard Gilder Graduate School at the AMNH.

Appendix A. Supplementary material

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ympev.2022.107576.

References

Abadi, S., Azouri, D., Pupko, T., Mayrose, I., 2019. Model selection may not be a mandatory step for phylogeny reconstruction. Nat. Commun. 10, 934. https://doi.org/10.1038/s41467-019-08822-w.

Abreu-Jr, E.F., Pavan, S.E., Tsuchiya, M.T.N., Wilson, D.E., Percequillo, A.R., Maldonado, J.E., 2020a. Spatiotemporal Diversification of Tree Squirrels: Is the South American Invasion and Speciation Really That Recent and Fast? Front. Ecol. Evol. 8, 1–14. https://doi.org/10.3389/fevo.2020.00230.

Abreu-Jr, E.F., Pavan, S.E., Tsuchiya, M.T.N., Wilson, D.E., Percequillo, A.R., Maldonado, J.E., 2020b. Museomics of tree squirrels: a dense taxon sampling of mitogenomes reveals hidden diversity, phenotypic convergence, and the need of a taxonomic overhaul. BMC Evol. Biol. 20, 1–25. https://doi.org/10.1186/s12862-020-01639-y.

Aghbolaghi, M.A., Ahmadzadeh, F., Kiabi, H.B., Keyghobadi, N., 2020. Evolutionary history of the Persian squirrel (*Sciurus anomalus*): It emerged on the Eurasian continent in the Miocene. Zool. Anz. 287, 17–24. https://doi.org/10.1016/j.jcz.2020.04.007.

Aghbolaghi, M.A., Ahmadzadeh, F., Kiabi, B., Keyghobadi, N., 2019. The permanent inhabitant of the oak trees: Phylogeography and genetic structure of the Persian squirrel (*Sciurus anomalus*). Biol. J. Linn. Soc. 127, 197–212. https://doi.org/10.1093/biolinnean/blz032.

Allen, J.A., 1915. Review of the south american Sciuridae. Bull. Am. Museum Nat. Hist. 34, 147–309.

Baca, S.M., Alexander, A., Gustafson, G.T., Short, A.E.Z., 2017. Ultraconserved elements show utility in phylogenetic inference of Adephaga (Coleoptera) and suggest paraphyly of 'Hydradephaga'. Syst. Entomol. 42, 786–795. https://doi.org/10.1111/syen.12244.

Blaimer, B.B., Lloyd, M.W., Guillory, W.X., Brady, S.G., 2016. Sequence capture and phylogenetic utility of genomic ultraconserved elements obtained from pinned insect specimens. PLoS ONE 11, 1–20. https://doi.org/10.1371/journal.pone.0161531.

Blom, M.P.K., Bragg, J.G., Potter, S., Moritz, C., 2017. Accounting for uncertainty in gene tree estimation: Summary-coalescent species tree inference in a challenging radiation of Australian lizards. Syst. Biol. 66, 352–366. https://doi.org/10.1093/sysbio/syw089.

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

Borowiec, M., 2019. Spruceup: fast and flexible identification, visualization, and removal of outliers from large multiple sequence alignments. J. Open Source Softw. 4, 1635. https://doi.org/10.21105/joss.01635.

Branstetter, M.G., Müller, A., Griswold, T.L., Orr, M.C., Zhu, C.D., 2021. Ultraconserved element phylogenomics and biogeography of the agriculturally important mason bee subgenus *Osmia* (*Osmia*). Syst. Entomol. 46, 453–472. https://doi.org/10.1111/syen.12470.

Bryson, R.W., Faircloth, B.C., Tsai, W.L.E., McCormack, J.E., Klicka, J., 2016. Target enrichment of thousands of ultraconserved elements sheds new light on early relationships within New World sparrows (Aves: Passerellidae). Auk 133, 451–458. https://doi.org/10.1642/auk-16-26.1.

Burgin, C.J., Colella, J.P., Kahn, P.L., Upham, N.S., 2018. How many species of mammals are there? J. Mammal. 99, 1–14. https://doi.org/10.1093/jmammal/gyx147.

Cai, L., Xi, Z., Lemmon, E.M., Lemmon, A.R., Mast, A., Buddenhagen, C.E., Liu, L., Davis, C.C., 2021. The Perfect Storm: Gene Tree Estimation Error, Incomplete Lineage Sorting, and Ancient Gene Flow Explain the Most Recalcitrant Ancient Angiosperm Clade, Malpighiales. Syst. Biol. 70, 491–507. https://doi.org/10.1093/sysbio/syaa083.

Calderón-Capote, M.C., Díaz-Nieto, J.F., López-Arévalo, H.F., 2015. Geographic distribution of the Pygmy Squirrel *Sciurillus pusillus* (É. Geoffroy-St.-Hilaire, 1803) (Rodentia: Sciuridae) in the northwestern Amazonia, southern Colombia. Check List 11, 1650. https://doi.org/10.15560/11.3.1650.

Card, D.C., Shapiro, B., Giribet, G., Moritz, C., Edwards, S.V., 2021. Museum Genomics. Annu. Rev. Genet. 55 https://doi.org/10.1146/annurev-genet-071719-020506.

Castañeda-Rico, S., León-Paniagua, L., Edwards, C.W., Maldonado, J.E., 2020. Ancient DNA From Museum Specimens and Next Generation Sequencing Help Resolve the Controversial Evolutionary History of the Critically Endangered Puebla Deer Mouse. Front. Ecol. Evol. 8, 1–18. https://doi.org/10.3389/fevo.2020.00094.

Castresana, J., 2000. Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. Mol. Biol. Evol. 17, 540–552. https://doi.org/10.1093/oxfordjournals.molbev.a026334.

Chan, K.O., Hutter, C.R., Wood, P.L., Grismer, L.L., Das, I., Brown, R.M., 2020. Gene flow creates a mirage of cryptic species in a Southeast Asian spotted stream frog complex. Mol. Ecol. 29, 3970–3987.

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

Chou, J., Gupta, A., Yaduvanshi, S., Davidson, R., Nute, M., Mirarab, S., Warnow, T., 2015. A comparative study of SVDquartets and other coalescent-based species tree estimation methods. BMC Genomics 16, S2. https://doi.org/10.1186/1471-2164-16-S10-S2.

Crawford, N.G., Faircloth, B.C., McCormack, J.E., Brumfield, R.T., Winker, K., Glenn, T. C., 2012. More than 1000 ultraconserved elements provide evidence that turtles are the sister group of archosaurs. Biol. Lett. 8, 783–786. https://doi.org/10.1098/rsbl.2012.0331.

D'Elía, G., Fabre, P.H., Lessa, E.P., 2019. Rodent systematics in an age of discovery: Recent advances and prospects. J. Mammal. 100, 852–871. https://doi.org/10.1093/jmammal/gyy179.

Derkarabetian, S., Benavides, L.R., Giribet, G., 2019. Sequence capture phylogenomics of historical ethanol-preserved museum specimens: Unlocking the rest of the vault. Mol. Ecol. Resour. 19, 1531–1544. https://doi.org/10.1111/1755-0998.13072.

Dornburg, A., Su, Z., Townsend, J.P., 2019. Optimal Rates for Phylogenetic Inference and Experimental Design in the Era of Genome-Scale Data Sets. Syst. Biol. 68, 145–156. https://doi.org/10.1093/sysbio/syy047.

Esselstyn, J.A., Oliveros, C.H., Swanson, M.T., Faircloth, B.C., 2017. Investigating difficult nodes in the placental mammal tree with expanded taxon sampling and thousands of ultraconserved elements. Genome Biol. Evol. 9, 2308–2321. https://doi.org/10.1093/gbe/evx168.

Fabre, P.-H., Hautier, L., Dimitrov, D., Douzery, E.J., 2012. A glimpse on the pattern of rodent diversication: a phylogenetic approach. BMC Evol. Biol. 12, 88. https://doi.org/10.1186/1471-2148-12-88.

Faircloth, B.C., 2016. PHYLUCE is a software package for the analysis of conserved genomic loci. Bioinformatics 32, 786–788. https://doi.org/10.1093/bioinformatics/btv646.

Faircloth, B.C., 2013. Illumiprocessor: a trimmomatic wrapper for parallel adapter and quality trimming. https://doi.org/10.6079/J9ILL.

Faircloth, B.C., McCormack, J.E., Crawford, N.G., Harvey, M.G., Brumfield, R.T., Glenn, T.C., 2012. Ultraconserved elements anchor thousands of genetic markers spanning multiple evolutionary timescales. Syst. Biol. 61, 717–726. https://doi.org/10.1093/sysbio/sys004.

Faircloth, B.C., Sorenson, L., Santini, F., Alfaro, M.E., 2013. A Phylogenomic Perspective on the Radiation of Ray-Finned Fishes Based upon Targeted Sequencing of Ultraconserved Elements (UCEs). PLoS ONE 8, 1–7. https://doi.org/10.1371/journal.pone.0065923.

Grabherr, M.G., Haas, B.J., Yassour, M., Levin, J.Z., Thompson, D.A., Amit, I., Adiconis, X., Fan, L., Raychowdhury, R., Zeng, Q., Chen, Z., Mauceli, E., Hacohen, N., Gnirke, A., Rhind, N., di Palma, F., Birren, B.W., Nusbaum, C., Lindblad-Toh, K., Friedman, N., Regev, A., 2011. Trinity: reconstructing a full-length transcriptome without a genome from RNA-Seq data. Nat. Biotechnol. 29, 644–652. https://doi.org/10.1038/nbt.1883.

Hahn, M.W., Nakhleh, L., 2016. Irrational exuberance for resolved species trees. Evolution 70, 7–17. https://doi.org/10.1111/evo.12832.

Hawkins, M.T.R., Leonard, J.A., Helgen, K.M., McDonough, M.M., Rockwood, L.L., Maldonado, J.E., 2016. Evolutionary history of endemic Sulawesi squirrels constructed from UCEs and mitogenomes sequenced from museum specimens. BMC Evol. Biol. 16, 1–16. https://doi.org/10.1186/s12862-016-0650-z.

Heintzman, P.D., Elias, S.A., Moore, K., Paszkiewicz, K., Barnes, I., 2014. Characterizing DNA preservation in degraded specimens of *Amara alpina* (Carabidae: Coleoptera). Mol. Ecol. Resour. 14, 606–615. https://doi.org/10.1111/1755-0998.12205.

Hoang, D.T., Chernomor, O., von Haeseler, A., Minh, B.Q., Vinh, L.S., 2018. UFBoot2: Improving the Ultrafast Bootstrap Approximation. Mol. Biol. Evol. 35, 518–522. https://doi.org/10.1093/molbev/msx281.

Hope, A.G., Malaney, J.L., Bell, K.C., Salazar-Miralles, F., Chavez, A.S., Barber, B.R., Cook, J.A., 2016. Revision of widespread red squirrels (genus: *Tamiasciurus*) highlights the complexity of speciation within North American forests. Mol. Phylogenet. Evol. 100, 170–182. https://doi.org/10.1016/j.ympev.2016.04.014.

Hosner, P.A., Faircloth, B.C., Glenn, T.C., Braun, E.L., Kimball, R.T., 2016. Avoiding missing data biases in phylogenomic inference: An empirical study in the landfowl (Aves: Galliformes). Mol. Biol. Evol. 33, 1110–1125. https://doi.org/10.1093/molbev/msv347.

Huang, H., He, Q., Kubatko, L.S., Knowles, L.L., 2010. Sources of Error Inherent in Species-Tree Estimation: Impact of Mutational and Coalescent Effects on Accuracy and Implications for Choosing among Different Methods. Syst. Biol. 59, 573–583. https://doi.org/10.1093/sysbio/syq047.

Jombart, T., 2008. Adegenet: A R package for the multivariate analysis of genetic markers. Bioinformatics 24, 1403–1405. https://doi.org/10.1093/bioinformatics/btn129.

Katoh, K., Standley, D.M., 2013. MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. Mol. Biol. Evol. 30, 772–780. https://doi.org/10.1093/molbev/mst010.

Linkem, C.W., Minin, V.N., Leaché, A.D., 2016. Detecting the anomaly zone in species trees and evidence for a misleading signal in higher-level skink phylogeny (Squamata: Scincidae). Syst. Biol. 65, 465–477. https://doi.org/10.1093/sysbio/syw001.

Manthey, J.D., Campillo, L.C., Burns, K.J., Moyle, R.G., 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. https://doi.org/10.1093/sysbio/syw005.

McCormack, J.E., Tsai, W.L.E., Faircloth, B.C., 2016. Sequence capture of ultraconserved elements from bird museum specimens. Mol. Ecol. Resour. 16, 1189–1203. https://doi.org/10.1111/1755-0998.12466.

McDonough, M.M., Parker, L.D., Rotzel McInerney, N., Campana, M.G., Maldonado, J.E., 2018. Performance of commonly requested destructive museum samples for mammalian genomic studies. J. Mammal. 99, 789–802. https://doi.org/10.1093/jmammal/gyy080.

McLean, B.S., Bell, K.C., Allen, J.M., Helgen, K.M., Cook, J.A., 2019. Impacts of Inference Method and Data set Filtering on Phylogenomic Resolution in a Rapid Radiation of Ground Squirrels (Xerinae: Marmotini). Syst. Biol. 68, 298–316. https://doi.org/10.1093/sysbio/syy064.

McLean, B.S., Helgen, K.M., Goodwin, H.T., Cook, J.A., 2018. Trait-specific processes of convergence and conservatism shape ecomorphological evolution in ground-dwelling squirrels. Evolution (N. Y.) 72, 473–489. https://doi.org/10.1111/evo.13422.

Mercer, J.M., Roth, V.L., 2003. The Effects of Cenozoic Global Change on Squirrel Phylogeny. Science 299, 1568–1572. https://doi.org/10.1126/science.1079705.

Miller, M.A., Pfeiffer, W., Schwartz, T., 2010. Creating the CIPRES Science Gateway for inference of large phylogenetic trees. In: 2010 Gateway Computing Environments Workshop (GCE). IEEE, pp. 1–8. https://doi.org/10.1109/GCE.2010.5676129.

Minh, B.Q., Schmidt, H.A., Chernomor, O., Schrempf, D., Woodhams, M.D., von Haeseler, A., Lanfear, R., 2020. IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the Genomic Era. Mol. Biol. Evol. 37, 1530–1534. https://doi.org/10.1093/molbev/msaa015.

Mirarab, S., Bayzid, M.S., Warnow, T., 2016. Evaluating summarymethods for multilocus species tree estimation in the presence of incomplete lineage sorting. Syst. Biol. 65, 366–380. https://doi.org/10.1093/sysbio/syu063.

Moore, J.C., 1959. Relationships among the living squirrels of the Sciurinae. Bull. Am. Museum Nat. Hist. 118, 157–206.

Nakamura, T., Yamada, K.D., Tomii, K., Katoh, K., 2018. Parallelization of MAFFT for large-scale multiple sequence alignments. Bioinformatics 34, 2490–2492. https://doi.org/10.1093/bioinformatics/bty121.

Oliveros, C.H., Field, D.J., Ksepka, D.T., Barker, F.K., Aleixo, A., Andersen, M.J., Alström, P., Benz, B.W., Braun, E.L., Braun, M.J., Bravo, G.A., Brumfield, R.T., Chesser, R.T., Claramunt, S., Cracraft, J., Cuervo, A.M., Derryberry, E.P., Glenn, T.C., Harvey, M.G., Hosner, P.A., Joseph, L., Kimball, R.T., Mack, A.L., Miskelly, C.M., Peterson, A.T., Robbins, M.B., Sheldon, F.H., Silveira, L.F., Smith, B.T., White, N.D., Moyle, R.G., Faircloth, B.C., 2019. Earth history and the passerine superradiation. Proc. Natl. Acad. Sci. 116, 7916–7925. https://doi.org/10.1073/pnas.1813206116.

Oshida, T., Arslan, A., Noda, M., 2009. Phylogenetic relationships among the old world *Sciurus* squirrels. Folia Zool. 58, 14–25.

Oshida, T., Masuda, R., 2000. Phylogeny and Zoogeography of Six Squirrel Species of the Genus *Sciurus* (Mammalia, Rodentia), Inferred from Cytochrome b Gene Sequences. Zoolog. Sci. 17, 405–409. https://doi.org/10.2108/zsj.17.405.

Parada, A., Hanson, J., D'Eiía, G., 2021. Ultraconserved Elements Improve the Resolution of Difficult Nodes within the Rapid Radiation of Neotropical Sigmodontine Rodents (Cricetidae: Sigmodontinae). Syst. Biol. 70, 1090–1100. https://doi.org/10.1093/sysbio/syab023.

Pečnerová, P., Martínková, N., 2012. Evolutionary history of tree squirrels (Rodentia, Sciurini) based on multilocus phylogeny reconstruction. Zool. Scr. 41, 211–219. https://doi.org/10.1111/j.1463-6409.2011.00528.x.

Pečnerová, P., Moravec, J.C., Martínková, N., 2015. A skull might lie: Modeling ancestral ranges and diet from genes and shape of tree squirrels. Syst. Biol. 64, 1074–1088. https://doi.org/10.1093/sysbio/syv054.

Platt, R.N., Faircloth, B.C., Sullivan, K.A.M., Kieran, T.J., Glenn, T.C., Vandewege, M.W., Lee, T.E., Baker, R.J., Stevens, R.D., Ray, D.A., 2018. Conflicting Evolutionary Histories of the Mitochondrial and Nuclear Genomes in New World *Myotis* Bats. Syst. Biol. 67, 236–249. https://doi.org/10.1093/sysbio/syx070.

Portik, D.M., Wiens, J.J., 2021. Do Alignment and Trimming Methods Matter for Phylogenomic (UCE) Analyses? Syst. Biol. 70, 440–462. https://doi.org/10.1093/sysbio/syaa064.

R Core Team, 2016. R: a language and environment for statistical computing. Vienna: R Foundation for Statistical Computing. Available at: http://www.R-project.org/.

Rocha, R.G., Leite, Y.L.R., Costa, L.P., Rojas, D., 2016. Independent reversals to terrestriality in squirrels (Rodentia: Sciuridae) support ecologically mediated modes of adaptation. J. Evol. Biol. 29, 2471–2479. https://doi.org/10.1111/jeb.12975.

Rohland, N., Reich, D., 2012. Cost-effective, high-throughput DNA sequencing libraries for multiplexed target capture. Genome Res. 22, 939–946. https://doi.org/10.1101/gr.128124.111.

Roth, V.L., Mercer, J.M., 2008. Differing rates of macroevolutionary diversification in arboreal squirrels. Curr. Sci. 95, 857–861.

Ruane, S., Austin, C.C., 2017. Phylogenomics using formalin-fixed and 100+ year-old intractable natural history specimens. Mol. Ecol. Resour. 17, 1003–1008. https://doi.org/10.1111/1755-0998.12655.

Sarkar, D., 2008. Lattice. New York, NY: Springer New York. https://doi.org/10.1007/978-0-387-75969-2.

Schliep, K.P., 2011. phangorn: Phylogenetic analysis in R. Bioinformatics 27, 592–593. https://doi.org/10.1093/bioinformatics/btq706.

Sharma, P.P., Fernández, R., Esposito, L.A., González-Santillán, E., Monod, L., 2015. Phylogenomic resolution of scorpions reveals multilevel discordance with morphological phylogenetic signal. Proc. R. Soc. B Biol. Sci. 282, 20142953. https://doi.org/10.1098/rspb.2014.2953.

Sheffield, N.C., 2013. The Interaction between Base Compositional Heterogeneity and Among-Site Rate Variation in Models of Molecular Evolution. ISRN Evol. Biol. 2013, 1–8. https://doi.org/10.5402/2013/391561.

Simmons, M.P., Gatesy, J., 2021. Collapsing dubiously resolved gene-tree branches in phylogenomic coalescent analyses. Mol. Phylogenet. Evol. 158, 107092. https://doi.org/10.1016/j.ympev.2021.107092.

Simmons, M.P., Sloan, D.B., Gatesy, J., 2016. The effects of subsampling gene trees on coalescent methods applied to ancient divergences. Mol. Phylogenet. Evol. 97, 76–89. https://doi.org/10.1016/j.ympev.2015.12.013.

Smith, B.T., Harvey, M.G., Faircloth, B.C., Glenn, T.C., Brumfield, R.T., 2014. Target capture and massively parallel sequencing of ultraconserved elements for comparative studies at shallow evolutionary time scales. Syst. Biol. 63, 83–95. https://doi.org/10.1093/sysbio/syt061.

Stamatakis, A., 2014. RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics 30, 1312–1313. https://doi.org/10.1093/bioinformatics/btu033.

Steppan, S.J., Storz, B.L., Hoffmann, R.S., 2004. Nuclear DNA phylogeny of the squirrels (Mammalia: Rodentia) and the evolution of arboreality from c-myc and RAG1. Mol. Phylogenet. Evol. 30, 703–719. https://doi.org/10.1016/S1055-7903(03)00204-5.

Streicher, J.W., Schulte, J.A., Wiens, J.J., 2016. How Should Genes and Taxa be Sampled for Phylogenomic Analyses with Missing Data? An Empirical Study in Iguanian Lizards. Syst. Biol. 65, 128–145. https://doi.org/10.1093/sysbio/syv058.

Streicher, J.W., Wiens, J.J., 2017. Phylogenomic analyses of more than 4000 nuclear loci resolve the origin of snakes among lizard families. Biol. Lett. 13 https://doi.org/10.1098/rsbl.2017.0393.

Swanson, M.T., Oliveros, C.H., Esselstyn, J.A., 2019. A phylogenomic rodent tree reveals the repeated evolution of masseter architectures. Proc. R. Soc. B Biol. Sci. 286 https://doi.org/10.1098/rspb.2019.0672.

Swofford, D., 2003. PAUP*. Phylogenetic Analysis Using Parsimony (*and Other Methods). Version 4. Sinauer Associates, Sunderland.

Talavera, G., Castresana, J., 2007. Improvement of phylogenies after removing divergent and ambiguously aligned blocks from protein sequence alignments. Syst. Biol. 56, 564–577. https://doi.org/10.1080/10635150701472164.

Tan, G., Muffato, M., Ledergerber, C., Herrero, J., Goldman, N., Gil, M., Dessimoz, C., 2015. Current methods for automated filtering of multiple sequence alignments frequently worsen single-gene phylogenetic inference. Syst. Biol. 64, 778–791. https://doi.org/10.1093/sysbio/syv033.

Tchaicka, L., de Freitas, T.R.O., Bager, A., Vidal, S.L., Lucherini, M., Iriarte, A., Novaro, A., Geffen, E., Garcez, F.S., Johnson, W.E., Wayne, R.K., Eizirik, E., 2016. Molecular assessment of the phylogeny and biogeography of a recently diversified endemic group of South American canids (Mammalia: Carnivora: Canidae). Genet. Mol. Biol. 39, 442–451. https://doi.org/10.1590/1678-4685-GMB-2015-0189.

Thorington, R.W., Koprowski, J.L., Steele, M.A., Whatton, J., 2012. Squirrels of the World. The Johns Hopkins University Press, Baltimore.

Trigo, T.C., Schneider, A., De Oliveira, T.G., Lehugeur, L.M., Silveira, L., Freitas, T.R.O., Eizirik, E., 2013. Molecular data reveal complex hybridization and a cryptic species of Neotropical wild cat. Curr. Biol. 23, 2528–2533. https://doi.org/10.1016/j.cub.2013.10.046.

Upham, N.S., Esselstyn, J.A., Jetz, W., 2019. Inferring the mammal tree: Species-level sets of phylogenies for questions in ecology, evolution, and conservation. PLoS Biol. 17, e3000494. https://doi.org/10.1371/journal.pbio.3000494.

Villalobos, F., Cervantes-Reza, F., 2007. Phylogenetic relationships of Mesoamerican species of the genus *Sciurus* (Rodentia: Sciuridae). Zootaxa 40, 31–40. https://doi.org/10.11646/zootaxa.1525.1.3.

Villalobos, F., Gutierrez-Espeleta, G., 2014. Mesoamerican tree squirrels evolution (Rodentia: Sciuridae): a molecular phylogenetic analysis. Rev. Biol. Trop. 62, 649–657. https://doi.org/10.15517/rbt.v62i2.10614.

Vivo, M., Carmignotto, A.P., 2015. Family Sciuridae G. Fischer, 1817. In: Patton, J.L., Pardiñas, U.F.J., D'Elía, G. (Eds.), Mammals of South America, vol. 2, Rodents. The University of Chicago Press, Chicago and London, pp. 1–48.

Wickham, H., 2016. ggplot2, Use R! Springer International Publishing, Cham. https://doi.org/10.1007/978-3-319-24277-4.

Zelditch, M.L., Li, J., Tran, L.A.P., Swiderski, D.L., 2015. Relationships of diversity, disparity, and their evolutionary rates in squirrels (Sciuridae). Evolution (N. Y.) 69, 1284–1300. https://doi.org/10.1111/evo.12642.

Zhang, C., Rabiee, M., Sayyari, E., Mirarab, S., 2018. ASTRAL-III: Polynomial time species tree reconstruction from partially resolved gene trees. BMC Bioinf. 19, 15–30. https://doi.org/10.1186/s12859-018-2129-y.