Syst. Biol. 72(1):62–77, 2023

https://doi.org/10.1093/sysbio/syac076
Advance Access Publication December 6, 2022

Evaluating the Impact of Anatomical Partitioning on Summary Topologies Obtained with Bayesian Phylogenetic Analyses of Morphological Data

DANIEL M. CASALI[1,2,3,*], FELIPE V. FREITAS[3] AND FERNANDO A. PERINI[1,2]

[1]*Departamento de Zoologia, Universidade Federal de Minas Gerais, Belo Horizonte 31270-901, Brazil*
[2]*Pós-Graduacão em Zoologia, Universidade Federal de Minas Gerais, Av. Antônio Carlos, 6627, Belo Horizonte 31270-901, Brazil*
[3]*Departamento de Biologia, Faculdade de Filosofia, Ciências e Letras, Universidade de São Paulo, Ribeirão Preto 14040-901, Brazil*
[*]*Correspondence to be sent to: Departamento de Zoologia, Universidade Federal de Minas Gerais, Belo Horizonte 31270-901, Brazil; E-mail: daniel_casali@yahoo.com.br.*

Received 3 December 2021; reviews returned 30 October 2022; accepted 30 November 2022
Associate Editor: Stephen Smith

Abstract.—Morphological data are a fundamental source of evidence to reconstruct the Tree of Life, and Bayesian phylogenetic methods are increasingly being used for this task. Bayesian phylogenetic analyses require the use of evolutionary models, which have been intensively studied in the past few years, with significant improvements to our knowledge. Notwithstanding, a systematic evaluation of the performance of partitioned models for morphological data has never been performed. Here we evaluate the influence of partitioned models, defined by anatomical criteria, on the precision and accuracy of summary tree topologies considering the effects of model misspecification. We simulated datasets using partitioning schemes, trees, and other properties obtained from two empirical datasets, and conducted Bayesian phylogenetic analyses. Additionally, we reanalyzed 32 empirical datasets for different groups of vertebrates, applying unpartitioned and partitioned models, and, as a focused study case, we reanalyzed a dataset including living and fossil armadillos, testing alternative partitioning hypotheses based on functional and ontogenetic modules. We found that, in general, partitioning by anatomy has little influence on summary topologies analyzed under alternative partitioning schemes with a varying number of partitions. Nevertheless, models with unlinked branch lengths, which account for heterotachy across partitions, improve topological precision at the cost of reducing accuracy. In some instances, more complex partitioning schemes led to topological changes, as tested for armadillos, mostly associated with models with unlinked branch lengths. We compare our results with other empirical evaluations of morphological data and those from empirical and simulation studies of the partitioning of molecular data, considering the adequacy of anatomical partitioning relative to alternative methods of partitioning morphological datasets. [Evolutionary rates; heterogeneity; morphology; Mk model; partition; topology.]

Morphological data are a fundamental source of evidence to reconstruct the Tree of Life. In most cases, morphology is the only kind of data available for phylogenetic inference of fossil taxa (Wiens 2004). Morphology can also complement molecular data in inferences of the phylogeny and timescale of living and extinct organisms (Lee and Palci 2015). Traditionally, morphological datasets have been analyzed with the maximum parsimony criterion, but recently, model-based phylogenetic analyses, in particular Bayesian methods, have become widely available (Wright 2019; Wright and Lloyd 2020). In parallel with the growing use of probabilistic approaches, there has been an increasing number of studies exploring the performance of alternative phylogenetic methods using both simulated and empirical morphological data (Wright and Hillis 2014; O'Reilly et al. 2016, 2018; Puttick et al. 2017, 2019; Goloboff et al. 2018b; Schrago et al. 2018; Goloboff and Arias 2019; Smith 2019a).

Bayesian phylogenetic analyses of discrete morphology typically use the Mk model (Lewis 2001), which is essentially the Jukes–Cantor model of nucleotide substitution (Jukes and Cantor 1969) generalized for any number of states. Since its original proposal, extensions of that model have been presented in order to account for the heterogeneity in evolutionary patterns present in morphological datasets, like allowing for unequal character state frequencies, nonstationarity, or the use of alternative distributions to model among-character rate variation (Harrison and Larsson 2015; Klopfstein et al. 2015; Wright et al. 2016; Pyron 2017).

Another important way to account for data heterogeneity in phylogenetic analyses is through partitioned models, in which subsets of the data have their parameters estimated independently from those of other subsets (Brown and Lemmon 2007; Lanfear et al. 2012). Parameters such as the substitution model exchangeability rates and nucleotide frequencies, distribution of among-character rate variation, and branch lengths can be independently inferred for each partition. The performance of partitioned models applied to molecular datasets is well explored and understood (Brown and Lemmon 2007; Kainer and Lanfear 2015; Duchêne et al. 2020). More recently, data partitioning began to be applied in studies using empirical morphological data, evaluating alternative partitioning criteria, including partitioning by anatomy, homoplasy, evolutionary rates, number of character states, neomorphic × transformational characters, and clusters in a morphospace (Clarke and Middleton 2008; Close et al. 2015; Felsinger 2019; Rosa et al. 2019; Simões and

Pierce 2021; Casali et al. 2022). These studies, however, evaluated partitioned models only for a single or a few datasets.

Here we evaluate the performance of data partitioning according to the anatomical criterion (Clarke and Middleton 2008; Tarasov and Génier 2015; Porto et al. 2021). This criterion considers that subsets of characters associated with distinct anatomical modules evolve at different rates (Tarasov and Génier 2015). These modules can be defined according to hypotheses of character's structural, functional, or developmental integration (Clarke and Middleton 2008). Usually, those partitions are defined for general and localized body regions like cranial × postcranial skeleton (Varela et al. 2019) or internal × external anatomy (Porto et al. 2021), but fine-grained hypotheses of partitioning have also been evaluated (Tarasov and Génier 2015; Casali et al. 2022). Distinctly from other partitioning criteria, especially those applied to molecular data, anatomical partitioning requires specific knowledge from researchers studying the morphology of a given taxonomic group as anatomy varies considerably more than molecules when distantly related groups are compared (e.g., angiosperms and mammals).

Although studying the performance of models using empirical datasets ensures a greater degree of realism, the use of simulated data allows the assessment of the accuracy of a given method or model (Hillis 1995), something not feasible to do using empirical datasets. To the best of our knowledge, no study to date evaluated the performance of partitioned models for morphological data using simulations.

We applied anatomically partitioned models for simulated morphological datasets in a Bayesian framework, and evaluated their impact on the precision and accuracy of the estimated summary topologies, considering the effects of using the correct partitioning scheme used to simulate the data, and alternative partitioning schemes, to explore the influence of partitioning mismodeling. We also reanalyzed 32 vertebrate empirical datasets originally proposed to be composed of two anatomical partitions—cranial and postcranial (Mounce et al. 2016)—considering unpartitioned and partitioned models. Finally, we also conducted a detailed evaluation of alternative partitioning hypotheses based on functional and ontogenetic information for a dataset of living and fossil armadillos (Barasoain et al. 2021).

We show that, in general, partitioning data by anatomy has little effect on the precision and accuracy of inferred topologies when branches are linked across partitions, whereas models with unlinked branch lengths across anatomical partitions produce more precise but less accurate consensus topologies. We also discuss the adequacy of anatomical partitioning relative to alternative methods of partitioning morphological data, and compare it to what we know about the performance of partitioned models for molecular datasets.

Material and Methods

Simulated Data

Simulations.—To simulate partitioned morphological datasets, we require a partitioning scheme, a reference tree (topology and branch lengths), and a few properties from empirical datasets (Fig. 1). Those were obtained from two studies (Clarke and Middleton 2008; Porto et al. 2021), in which the authors applied anatomical partitioning in Bayesian phylogenetic inference. In both studies, models using anatomical partitioning presented a better fit to the data than unpartitioned models, according to Bayes factors. Although the original Bayes factor comparisons for the first study were conducted with an unreliable estimator of marginal likelihoods, the harmonic mean (Xie et al. 2011), a recent study confirmed the results using stepping-stone sampling (Rosa et al. 2019).

The first study, by Clarke and Middleton (2008), was the first to evaluate the use of anatomical partitioning in Bayesian phylogenetics. This study investigated the phylogeny of birds (Dinosauria: Avialae) and included 25 taxa, most of them extinct. It explored alternative anatomical partitioning schemes with two, three, and four partitions for the dataset, which will be henceforth referred to as dataset A. The second study, by Porto et al. (2021), investigated anatomical partitioning by studying the phylogeny of corbiculate bees (Hymenoptera: Apidae) and several outgroups, including 50 extant taxa. Although this study mainly investigated informational dissonance among partitions, it also performed phylogenetic analyses and Bayes factor comparisons. It was also chosen because it contains a partitioning scheme with a greater number of partitions than those applied to dataset A and is entirely composed of living taxa, contrasting with the first dataset. Anatomical partitioning schemes with two and seven partitions were proposed for this dataset in the original study, which will henceforth be referred to as dataset B.

To obtain reference trees for datasets A and B, they were reanalyzed with MrBayes 3.2.7a (Ronquist et al. 2012), replicating the settings of character ordering (or lack thereof) as used in the original analyses. For each dataset and considering their original anatomical partitioning schemes (Table 1), we applied three analytical approaches—(i) unpartitioned analyses, (ii) partitioned analyses with linked branch lengths, with among-partition rate variation accounted for by partition-specific rate multipliers, and (iii) partitioned analyses with unlinked branch lengths, which accounts for among-partition heterotachy (Marshall et al. 2006). We used the Mkv model of character evolution (Lewis 2001) and applied a discrete gamma distribution with four-rate categories to account for among-character rate variation (Yang 1996; Harrison and Larsson 2015), with unlinked estimates of shape (α) for each partition. The Monte Carlo Markov chain settings varied among

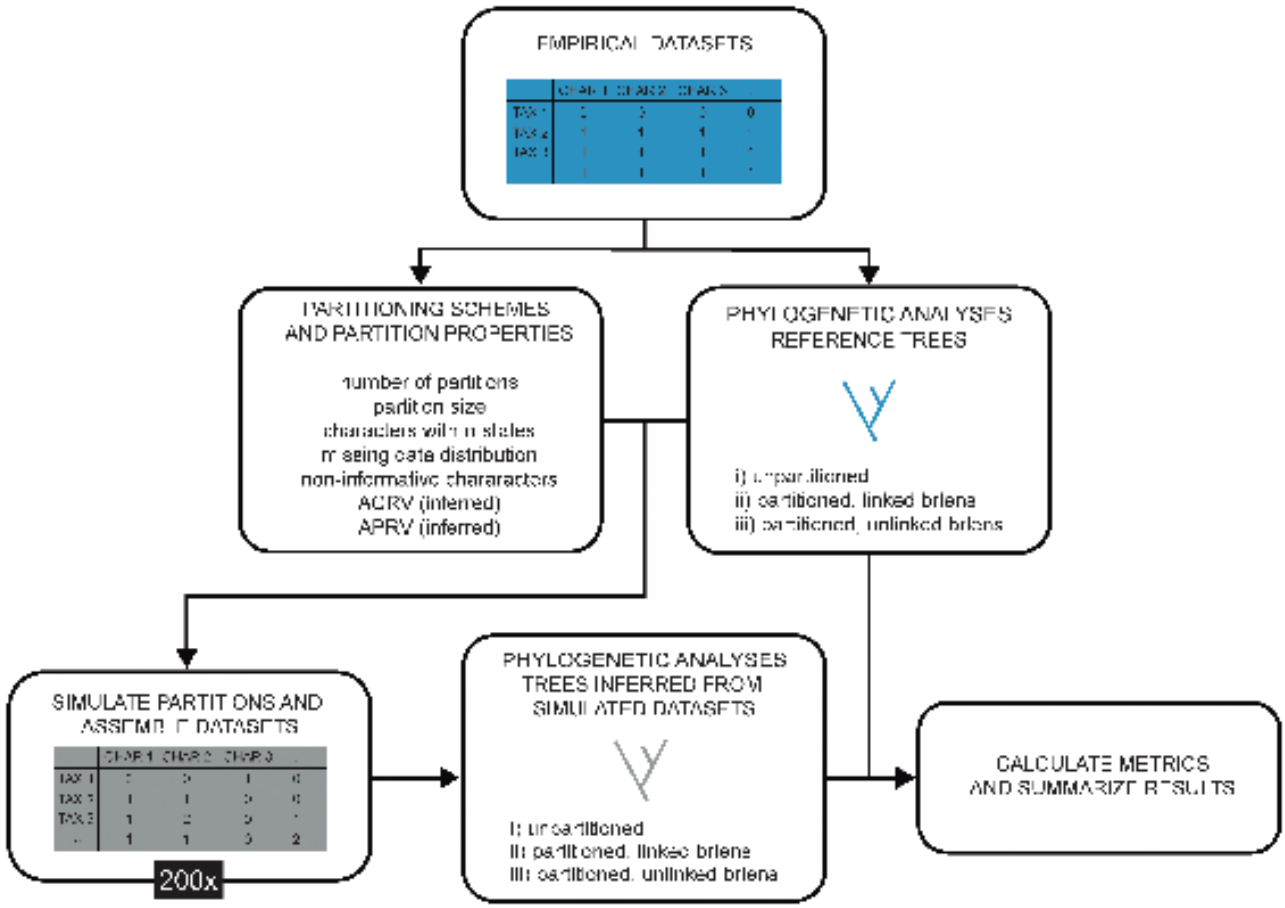

FIGURE 1. Graphical summary of the methodological steps used in the simulation and the analyses performed for simulated data. ACRV—Among-character rate variation. APRV—Among-partition rate variation.

TABLE 1. Properties of partitions considered for empirical datasets A and B, and their associated partitioning schemes

Dataset	Scheme	Partition	Characters	Non-info	Missing	Binary	Three st.	Four st.	Five st.	Six st.
A	s1	ALL	205	16	0.39	0.75	0.22	0.02	0.01	-
A	s4	CRANIAL	52	7	0.53	0.81	0.19	-	-	-
A	s4	AXIAL	19	0	0.37	0.58	-	0.37	-	0.05
A	s2, s3, s4	PECTORAL	83	3	0.34	0.81	0.14	0.02	0.02	-
A	s3, s4	PELVIC	51	6	0.31	0.65	0.31	0.04	-	-
A	s3	CRANIAL_AXIAL	71	7	0.49	0.75	0.14	0.10	-	0.01
A	s2	CRANIAL_AXIAL_PELVIC	122	13	0.42	0.70	0.21	0.07	-	0.01
B	s1	ALL	282	1	0.09	0.69	0.26	0.06	-	-
B	s2	EXT	181	0	0.06	0.71	0.25	0.04	-	-
B	s2	INT	101	1	0.16	0.64	0.27	0.09	-	-
B	s7	HD	42	0	0.05	0.74	0.24	0.02	-	-
B	s7	MP	52	0	0.08	0.79	0.17	0.04	-	-
B	s7	MS	57	0	0.03	0.68	0.26	0.05	-	-
B	s7	WG	16	0	0.09	0.56	0.44	-	-	-
B	s7	LG	49	0	0.07	0.49	0.37	0.14	-	-
B	s7	MT	11	0	0.20	0.82	0.18	-	-	-
B	s7	GN	55	1	0.21	0.75	0.20	0.05	-	-

Note. Characters—total number of characters. Non-info—number of noninformative characters. Missing—average proportion of missing cells per character. st. – proportion of characters for a given number of states.

EXT = external; INT = internal; HD = head; MP = mouthparts; MS = mesosoma; WG = wings; LG = legs; MT = metasoma; GN = genitalia.

analyses and are detailed in Supplementary Data S1. Trees were summarized as maximum compatibility trees (*contype* = *allcompat*). Convergence of continuous parameters and topologies was assessed with the function *analyze.rwty* in the package *rwty* (Warren et al. 2017) in R environment (R Core Team 2022). We visually inspected all trace plots and considered that analyses converged when individual runs achieved an effective

sample size (ESS) >200, an average standard deviation of split frequencies (ASDSF) <0.01, and a correlation of splits between runs ≥0.99. All inputs, outputs, and files used to assess convergence are available as Supplementary Data S1.

We considered seven simulation schemes for dataset A (s1—data unpartitioned, s2L—two partitions with linked branch lengths, s2U—two partitions with unlinked branch lengths, s3L—three partitions with linked branch lengths, s3U—three partitions with unlinked branch lengths, s4L—four partitions with linked branch lengths, and s4U—four partitions with unlinked branch lengths). For dataset B, five partitioning schemes were applied (s1—data unpartitioned, s2L—two partitions with linked branch lengths, s2U—two partitions with unlinked branch lengths, s7L—seven partitions with linked branch lengths, s7U—seven partitions with unlinked branch lengths).

Characters were simulated considering the topology and branch lengths (in units of substitution) present in the reference trees. In addition to partitioning schemes and trees, properties from the original partitions were also obtained, and applied in simulations—(i) partition size (i.e., number of characters per partition), (ii) number of noninformative characters, (iii) proportion of characters for each number of states (e.g., binary, with three states), (iv) proportion of missing cells per character, (v) distribution of among-character rate variation, and (vi) partition rate multipliers. Properties "i–iv" were obtained directly from the datasets, whereas properties "v–vi" have been estimated during the reanalyses of empirical datasets, with "vi" only applying to reference trees obtained with linked branch lengths. Unlinked empirical analyses return a separate set of branch lengths for each partition, which were applied during simulations of partitions evolving according to unliked branch lengths. To simulate partitions evolving according to linked branch lengths, branch lengths of the reference topology were multiplied by the empirical partition rate multiplier.

Characters in each partition were simulated in R using the package *dispRity* (Guillerme 2018). We used the function *sim.morpho* to generate variable characters, with the number of characters with a given number of states in each partition following the proportions obtained from empirical datasets (Table 1). After simulating each character, we checked if it was generated as a parsimony-informative character, and if not, the character was discarded and resimulated until this condition was satisfied. This procedure was repeated until we achieved the desired partition size, as informed by the empirical data (Table 1). Characters were simulated considering the Mkv+G model, with the rate/shape parameter of gamma distributions for each partition obtained from the reanalyses of the two empirical datasets.

Missing data are frequently present in morphological datasets, especially those including or exclusively composed of fossil taxa, with their distribution rarely being random (Prevosti and Chemisquy 2009). We attributed missing data to the simulated partitions by recoding cells to "?," sampling the number of missing cells per character from the empirical distribution of the respective partition, thus mimicking a nonrandom distribution of missing data. Whenever a character was turned noninformative due to this process, the missing data attribution was redone, ensuring that all modified characters were kept as parsimony informative. After that, we converted some characters to noninformative (all taxa coded as 0) in the proportion they were observed in that condition in empirical partitions (Table 1). We did not differentiate genuine missing data from inapplicable characters in this study since MrBayes consider them in the same manner. We generated 200 replications of each partition/dataset to account for the stochasticity in the character's simulation. The custom R scripts used for the simulations and the simulated datasets are available as Supplementary Data S2.

Phylogenetic analyses.—We analyzed each simulated dataset with all strategies of data partitioning (x) also used to simulate them (s), resulting in 49 combinations for dataset A and 25 for dataset B (Fig. 2). In that way, we were able to explore the influence of two types of modeling mismatch in the summary topologies—(i) partition number mismatch, with underpartitioned, correctly partitioned (i.e., matching the number of partitions that data were simulated and analyzed with) and overpartitioned models being considered, and (ii) branch length mismatch, considering analyses with linked and unlinked branch lengths across partitions. Analyses were performed with Bayesian inference in MrBayes, applying the Mkv+G model of character evolution, to be consistent with the conditions used to simulate the data. We initially set all analyses with two runs of 10M generations, with four chains each, sampling every 500th generation. The initial 25% of the samples were discarded as burn-in before summarizing continuous parameters and trees, with the latter being summarized into a maximum compatibility tree. Analyses were terminated before achieving the total number of generations when ASDSF reached values below 0.01. We set the *diagnfreq* to 5M generations to ensure a minimum sample size before checking ASDSF and finishing each analysis. Convergence was checked using ASDSF < 0.01, and ESS >100 for each run, using functions *mcmc* and *effectiveSize* in R package *coda* (Plummer et al. 2006). We also obtained 50% majority-rule consensus trees collapsing nodes with posterior probabilities <0.5 in the maximum compatibility tree in R, using the functions *read_annotated* of package *phylotate* (Beer and Beer 2019) and *collapseUnsupportedEdges* of the package *ips* (Heibl 2008). Summary trees and the files with the convergence assessment are available as Supplementary Data S2.

Topological similarity, precision, and accuracy.—We compared the summary topologies obtained from the analyses of each simulated dataset to the topology of the

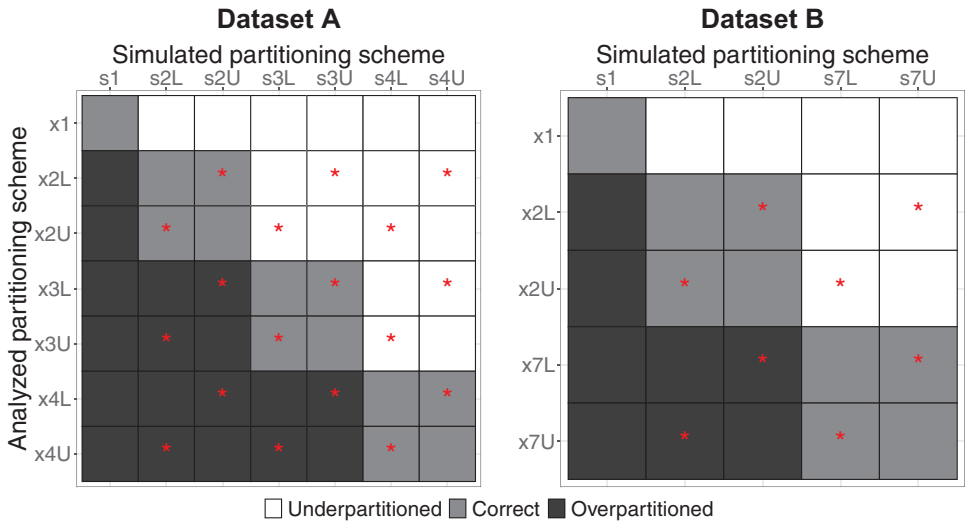

FIGURE 2. All combinations of simulated (s) and analyzed (x) partitioning schemes for datasets A and B, indicating underpartitioned, correct, and overpartitioned models. Asterisks indicate cases of branch length mismatch between the simulated and analyzed models.

reference tree used to simulate it (Fig. 1). Topological similarity was evaluated with the Mutual Clustering Information (MCI), an information-theoretic generalized Robinson–Foulds metric available in the package *TreeDist* (Smith 2020a, 2020b), and the Quartet Divergence metric (QD), available in the package *Quartet* (Smith 2019b, 2019a). Both metrics range from 0 (complete dissimilarity) to 1 (complete similarity). Nonetheless, these distance metrics are limited if we wish to make an independent assessment of the accuracy and precision of summary topologies that are not fully resolved, as is the case of the 50% majority-rule consensus topology, since it can conflate similarity due to correct resolution with that from lack of resolution (Keating et al. 2020). Thus, we also adopted two additional metrics. For precision, we calculated the proportion of resolved bipartitions/quartets (actual number of resolved/maximum possible number of resolved bipartitions/quartets). For accuracy, we calculated the proportion of correct bipartitions/quartets (correctly resolved/total number of resolved bipartitions/quartets). Bipartition and quartet status were calculated using the functions *SplitStatus* and *QuartetStatus*, respectively, in the package *Quartet* (Smith 2019b). For the purpose of this study, we are defining precision as the degree of resolution of a consensus topology. This is only a partial aspect of precision in Bayesian analysis, which can also be assessed, for example, by counting the number of unique topologies in the posterior sample (or in the 95% HPD), or more thoroughly, by inspecting the obtained tree space (Wright and Lloyd 2020; Smith 2022).

Before summarizing the results, we excluded all analyses that failed to converge in at least one of the inspected criteria (9% for dataset A and 13% for dataset B, Supplementary Data S2). Convergence issues were present in most groups (i.e., unique combinations of s and x) but were more prevalent in analyses with partitioning schemes with more partitions and/or with unlinked branch length, in which the number of parameters to be inferred is greater, with partitioning 7U of dataset B being the most affected.

We assessed the significance of the difference between pairs of distributions of metrics calculated for alternative partitioning schemes with a Wilcoxon rank-sum test (Wilcoxon 1945), with $\alpha = 0.05$, applying a Bonferroni correction to control for multiple comparisons (Bonferroni 1936). Pairwise comparisons between all analyzed partitioning schemes were considered (e.g., x1 to x2L, x1 to x2U) for each simulated partitioning scheme (e.g., s1, s2L, s2U). We expect that, if anatomical partitioning is an effective way to account for rate heterogeneity in morphological datasets, when simulated and analyzed partitioning schemes match (e.g., s3U and x3U), more similar, precise, and accurate topologies will be obtained, if compared to analyses not matching the simulated partitioning scheme. On the other hand, if distributions of metrics of topological similarity, accuracy, and precision end up being very similar for alternative partitioning schemes, anatomical partitioning can be regarded as of little importance for tree topology estimation. Additionally, a gain of precision accompanied by a decrease in accuracy can be considered an undesirable property of a given partitioning scheme, indicating a spurious resolution of the relationships among taxa. The opposite would be a positive feature, rendering the inference more conservative in not resolving the topology inaccurately.

Empirical Data

Datasets.—For a systematic evaluation of the performance of anatomical partitioning on empirical morphological data, we investigated several vertebrate datasets

previously analyzed by Mounce et al. (2016), reanalyzing them using partitioned models with Bayesian inference. Mounce et al. (2016) defined two anatomical partitions for all datasets, one for cranial and another for postcranial characters. When analyzed separately using maximum parsimony, in some instances, those two partitions produced significantly different topologies according to the incongruence relationship difference (IRD) test (Mounce et al. 2016). We selected all 32 datasets that satisfied that condition for one or both versions of the IRD test used in the original study to be reanalyzed here applying the same partitions proposed in Mounce et al. (2016). Supplementary Data S2 contains a list of those datasets and details of their partitions and other properties.

Additionally, we conducted a detailed case study for a dataset of Cingulata (Mammalia, Xenarthra) from the study of Barasoain et al. (2021), originally composed of 148 craniomandibular and carapace characters for 25 living and extinct armadillos, plus two outgroup taxa. We investigate seven alternative partitioning schemes: (i) unpartitioned; (ii) functional partition F1, which separates craniomandibular characters from those of the carapace (two partitions); (iii) functional partition F2, which further separates cranial characters according to three modules (oral/nasal, orbital/braincase, and basicranium/ear), and the carapace (four partitions); (iv) ontogenetic partition O1, in which character were segregated according to the developmental ossification pattern (dermal × endochondral, two partitions); (v) ontogenetic partition O2, according to the germ layer giving rise to the bones associated with the character (neural crest cells × mesoderm, two partitions); (vi) mixed partition M1, as O1, but with carapace characters allocated in a separate partition (three partitions); and (vii) mixed partition M2, as O2, but with carapace characters allocated in a separate partition (three partitions). The functional partitioning scheme F2 was defined following the mammalian modules recognized by Porto et al. (2009), considering the partial overlapping of modules observed for Cingulata. The ontogenetic association of mammalian cranial bones and cingulate carapace was obtained from reviews and recent studies (Novacek 1993; Noden and Trainor 2005; Kardong 2012; Koyabu et al. 2014; Maier and Ruf 2016; Krmpotic et al. 2021). Five of the 148 characters were removed from all the analyses since they could not be unambiguously associated with any partition due to our incomplete knowledge regarding the ontogeny of some bones (i.e., septomaxilla and entotympanic), or in cases in which the character is irreducibly associated with multiple bones with different ossification patterns or developmental origins (e.g., the sagittal crest, associated with frontals—originated from the neural crest, and with parietals—originated from the mesoderm). A list of the characters and their assignment to partitions is available in Supplementary Data S3.

Phylogenetic analyses.— The 32 vertebrate datasets were reanalyzed using Bayesian inference, with three approaches—(i) unpartitioned, (ii) partitioned, with linked branch lengths, and (iii) partitioned, with unlinked branch lengths. We applied the character ordering (or lack thereof) as in the original datasets and used the Mkv+G model. Analyses were performed in two runs, with four chains each, for 10M generations, sampling every 2000th. Some datasets demanded more generations to converge (20–25M). A burn-in of 25% was applied, and trees were summarized into maximum compatibility consensus trees. We also obtained a 50% majority-rule consensus in R, as reported for simulated data. Convergence was checked with *rwty*, as described for empirical datasets A and B analyses. Two of the 32 vertebrate datasets consistently failed to converge in unlinked analyses and were not considered in the results. All inputs, outputs, and files used to assess convergence are available in Supplementary Data S3.

The Cingulata dataset was also subjected to the same phylogenetic inference procedures described above for the vertebrate datasets, but with 5M generations, sampling every 500th. All inputs, outputs, and files used to assess convergence are available in Supplementary Data S4.

Topological precision and similarity.—The metrics used to evaluate precision in the analyses of simulated data were also applied to the vertebrate empirical datasets, comparing consensus tree resolution across alternative partitioning schemes. Also, the proportion of bipartitions/quartets resolved in the same way, and the topological similarity was computed. As for simulated data, significance was assessed with a Wilcoxon rank-sum test. For the Cingulata dataset, the topological similarity was evaluated for all pairs of partitioning schemes.

All R scripts used for calculating the metrics of similarity, accuracy, and precision; plot the results; and calculate summary and test statistics for simulated and empirical datasets are available as Supplementary Data.

Results

Simulated Data

General remarks.—Data simulated with unlinked branch lengths returned topologies more similar to reference topologies if compared to data simulated with linked branch lengths or unpartitioned. This pattern is also associated with the increase in the number of partitions applied in simulations with unlinked branch lengths. This pattern can be observed irrespective of the metric of similarity applied, collapsing regime (maximum compatibility tree—allcompat, 50% majority-rule consensus—halfcompat), or dataset (Figs. 3a and 4a, Supplementary Data S4 and S5). For simulations s3U and s4U of datasets A, and s7U of dataset B, very accurate topologies were recovered (also well resolved in the case of s7U), not showing differences associated

with the partitioning scheme used to analyze the data (Figs. 3b and 4b, Supplementary Data S4 and S5).

Topological similarity.—For both datasets and collapsing regimes, mismatches related to the number of partitions in simulated and analyzed conditions were of little consequence to topological similarity (Figs. 3a and 4a, Supplementary Data S4 and S5). For dataset A, when simulated and analyzed branch lengths condition matched, topologies tended to be equally or more similar to those of their respective reference tree (except in simulation s3L, Fig. 3a, Supplementary Data S4). On the other hand, for dataset B analyses with unlinked branch lengths mostly returned equally or more dissimilar topologies compared to those obtained with unpartitioned and linked models, irrespective of how branch lengths were simulated (Fig. 4a, Supplementary Data S5). Despite that, only allcompat topologies from analyses x7U of simulation s1 of dataset B showed significant differences compared to unpartitioned and linked partitioned analyses (Fig. 5, Supplementary Data S6 and S7).

Topological precision.—As for similarity, matching the number of partitions in simulated and analyzed conditions had a negligible impact on topological resolution (Figs. 3b and 4b, Supplementary Data S4 and S5). However, when the data were analyzed with unlinked branch lengths, more resolved topologies were obtained than in analyses with unpartitioned or linked models (Figs. 3b and 4b, Supplementary Data S4 and S5), with most of these differences being statistically significant (Fig. 5, Supplementary Data S6 and S7).

Topological accuracy.—As for similarity and precision, matching the number of partitions in simulated and analyzed conditions has not led to substantial changes in topological accuracy (Figs. 3b and 4b, Supplementary Data S4 and S5). In contrast to what was observed for

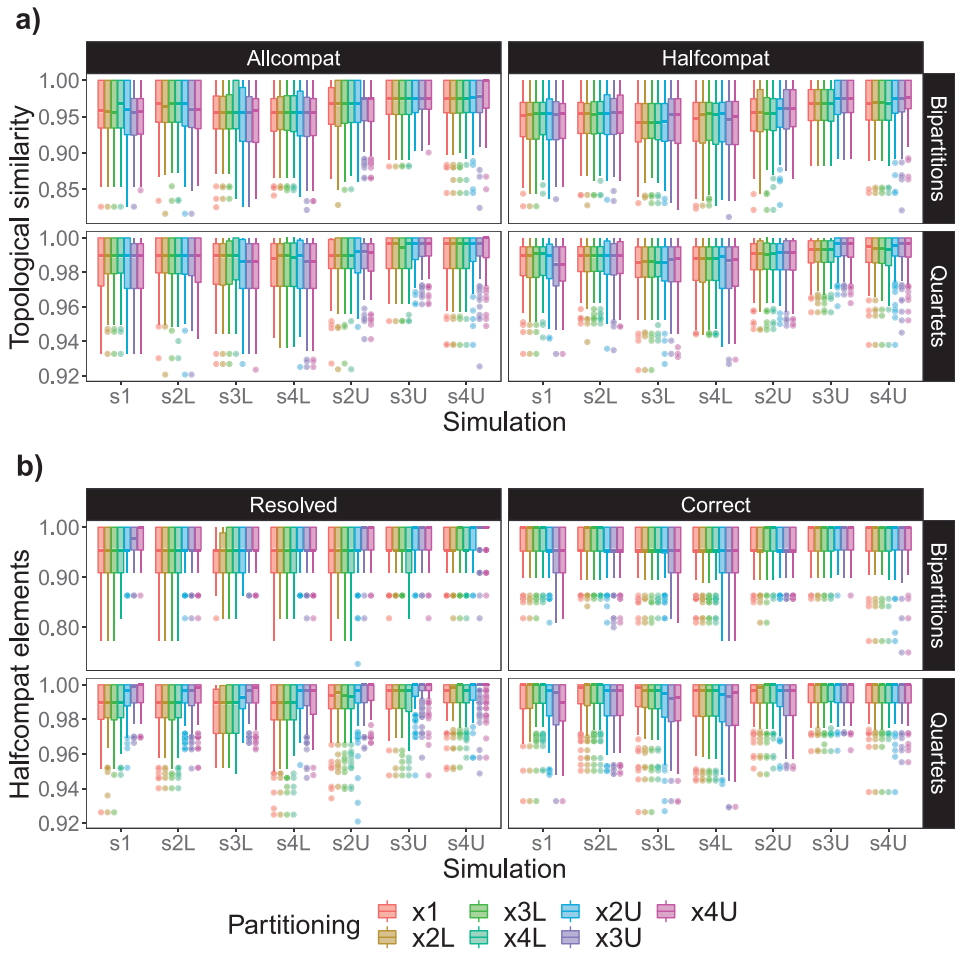

FIGURE 3. Topological similarity, precision, and accuracy evaluated by comparing topologies obtained by analyzing the simulated data (dataset A) to topologies of their respective reference trees. (a) Topological similarity, assessed with Mutual Cluster Information (bipartitions) and Quartet Divergence (quartets) for maximum compatibility (allcompat) and 50% majority-rule consensus trees (halfcompat). (b) Proportion of resolved and correct bipartitions/quartets in halfcompat trees. Boxplots summarize the metrics for all combinations of simulated (s) and analyzed (x) partitioning schemes.

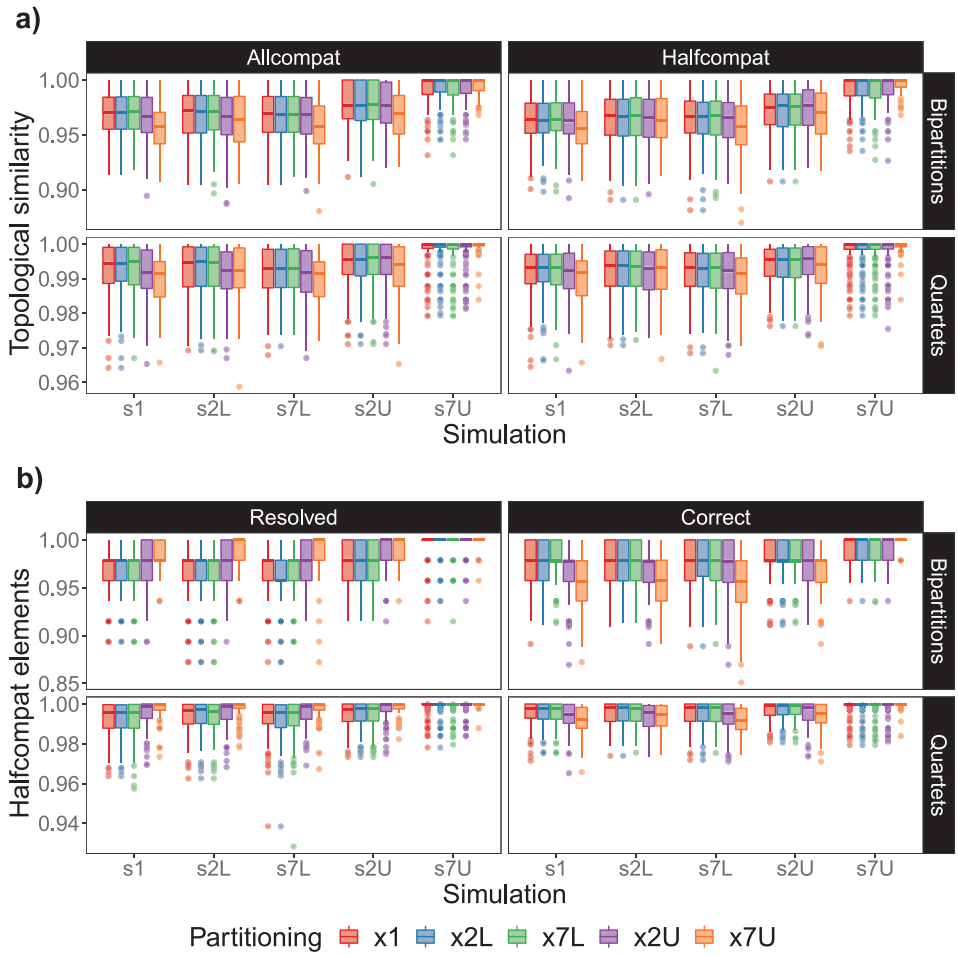

FIGURE 4. Topological similarity, precision, and accuracy evaluated by comparing topologies obtained by analyzing the simulated data (dataset B) to topologies of their respective reference trees. (a) Topological similarity, assessed with Mutual Cluster Information (bipartitions) and Quartet Divergence (quartets) for maximum compatibility (allcompat) and 50% majority-rule consensus trees (halfcompat). (b) Proportion of resolved and correct bipartitions/quartets in halfcompat trees. Boxplots summarize the metrics for all combinations of simulated (s) and analyzed (x) partitioning schemes.

precision, accuracy usually decreases in topologies resulting from analyses with unlinked branch lengths. On the other hand, topologies from partitioned analyses with linked branch lengths and unpartitioned analyses were at least equally, but frequently more accurate (Figs. 3b and 4b, Supplementary Data S4 and S5). Although these differences are apparent upon inspecting the median values in boxplots, they were only statistically significant in comparisons including analyses x7U of dataset B, which returned considerably less accurate topologies (Figs. 4 and 5, Supplementary Data S5–S7).

Empirical Data

Analyses of vertebrate empirical datasets showed results aligned with those observed for simulations. Topologies obtained with unpartitioned models were very similar to those obtained with partitioned models with linked branch lengths (median similarity >0.99, Fig. 6a, Supplementary Data S8), whereas models with unlinked branch lengths led to slightly less similar topologies relative to the other two treatments of branch lengths (median similarity ~0.93, Fig. 6a, Supplementary Data S8).

In five of eight pairwise comparisons, the similarity between topologies obtained with unpartitioned and linked models was significantly different from the similarity between topologies obtained with unpartitioned versus unlinked and from the similarity between topologies obtained with linked versus unlinked comparisons (Fig. 6a, Supplementary Data S9). However, when comparing both distributions of similarity including the unlinked model (i.e., unpartitioned vs. unlinked and linked vs. unlinked), they were not significantly different (Fig. 6a, Supplementary Data S9). The greater similarity between topologies obtained with unpartitioned and linked models was also reflected in the number of common bipartition/quartets, with topologies resulting from unpartitioned and linked analyses sharing a few more bipartitions/quartets than each of them shares with those topologies obtained from unlinked analyses

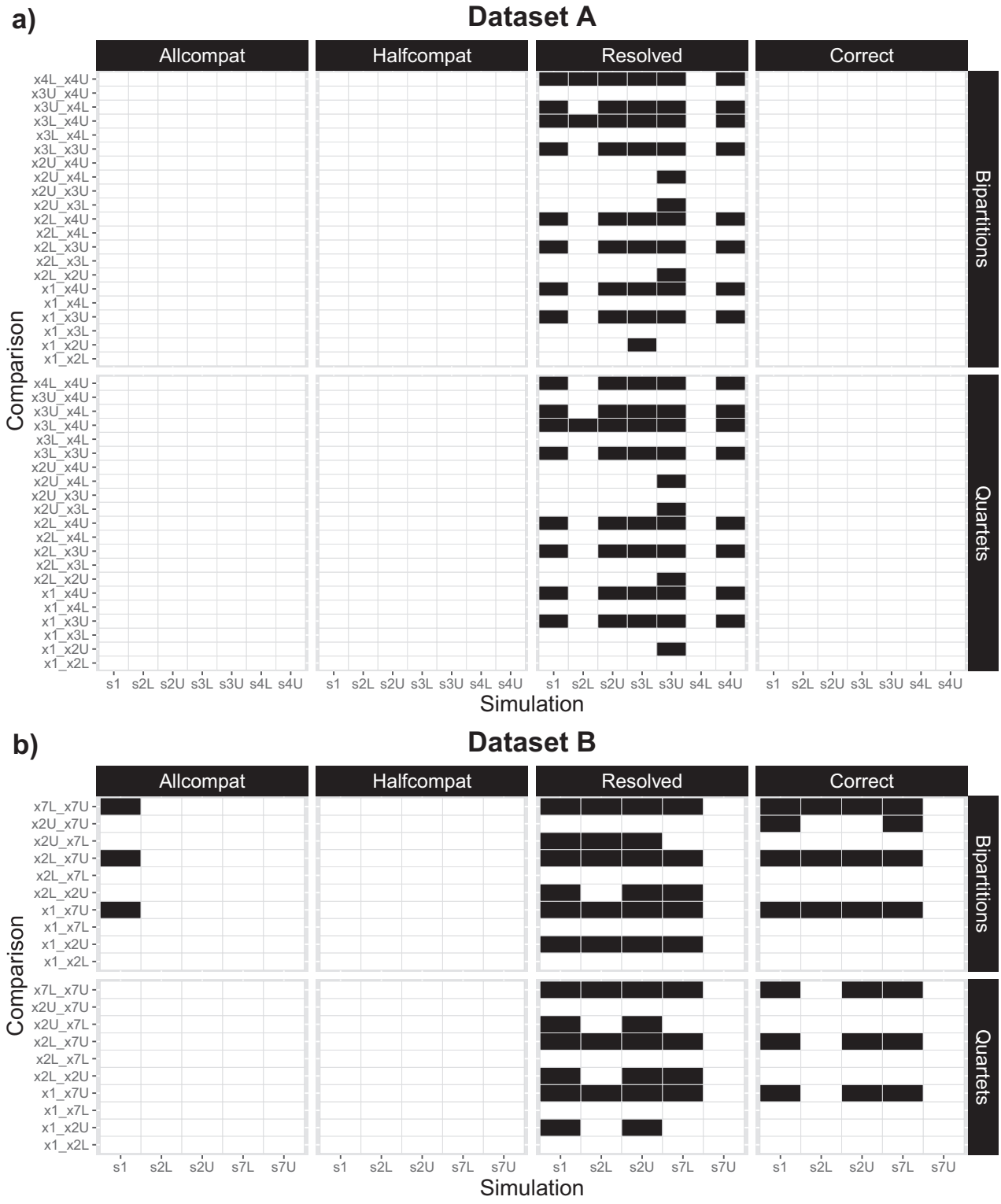

FIGURE 5. Summary of Wilcoxon rank-sum test for the significance of the difference between metrics calculated for alternative partitioning schemes (x) in each simulation condition (s). Significant differences are indicated in black. (a) Dataset A. (b) Dataset B. Metrics—topological similarity, assessed with Mutual Cluster Information (bipartitions), and Quartet Divergence (quartets) for maximum compatibility (allcompat) and 50% majority-rule consensus trees (halfcompat), and the proportion of resolved and correct bipartitions/quartets in halfcompat trees.

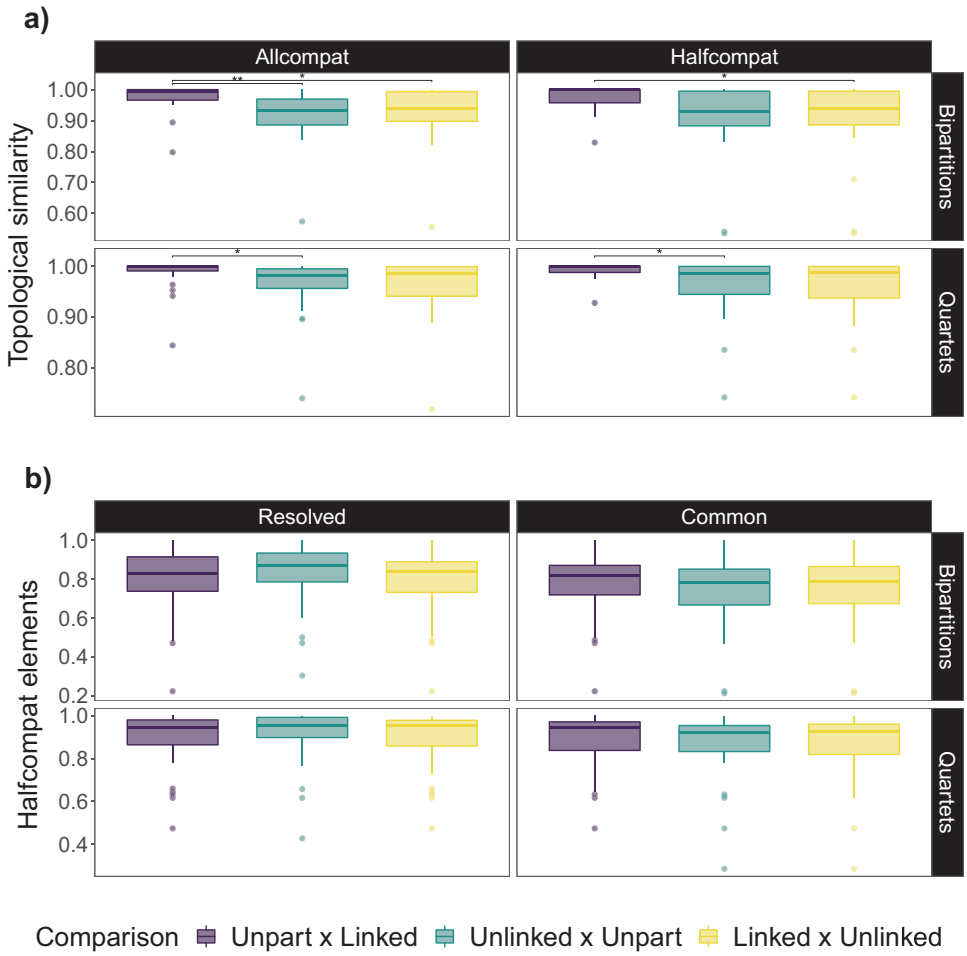

FIGURE 6. Topological similarity, precision, and the common number of bipartitions/quartets shared by topologies inferred under alternative partitioning schemes—unpartitioned (Unpart); partitioned, with linked branch lengths (Linked); and partitioned, with unlinked branch lengths (Unlinked). (a) Topological similarity, assessed with Mutual Cluster Information (bipartitions), and Quartet Divergence (quartets), for maximum compatibility (allcompat) and 50% majority-rule consensus trees (halfcompat). (b) Proportion of resolved and common bipartitions/quartets in halfcompat trees. The proportion of resolved bipartitions/quartets refers to individual models—Unpart, Unlinked, and Linked, respectively—and not to pairwise comparisons, as for the other metrics. Brackets indicate significant results.

(Fig. 6b, Supplementary Data S8). Nevertheless, none of the pairwise comparisons indicated statistically significant differences (Fig. 6b, Supplementary Data S8). As also observed for simulated data, partitioning schemes with unlinked branch lengths resulted in slightly better resolved halfcompat topologies, but these differences were not statistically significant either (Fig. 6b, Supplementary Data S8 and S9).

The topologies from the allcompat trees obtained from the analyses of the Cingulata mostly conformed to the general pattern observed for simulations and the vertebrate datasets. All unpartitioned and linked models returned identical topologies, whereas three of five unlinked models led to different topologies (Fig. 7). Nevertheless, halfcompat topologies differ in their degree of resolution, unrelated to the treatment of branch lengths (Table 2), with alternative partitioning hypotheses leading to slightly different topologies (Fig. 7, Supplementary Data S4).

Discussion

The Performance of Anatomical Partitioning

The use of adequate models is of paramount importance for statistical phylogenetics, and partitioned morphological models are only beginning to be investigated regarding their performance (Rosa et al. 2019; Casali et al. 2022). We present here the first simulation study of the performance of anatomical partitioning, as well as the first study of this kind for morphological partitioning in general. We also present the first systematic evaluation of anatomical partitioning applied to several morphological datasets in Bayesian phylogenetics. Previous studies that conducted a detailed evaluation of anatomical partitioned models for morphological datasets were limited to exploring only one or a few empirical datasets and focused mainly (though not exclusively) on model selection using Bayes factor (Tarasov and Génier 2015; Rosa et al. 2019;

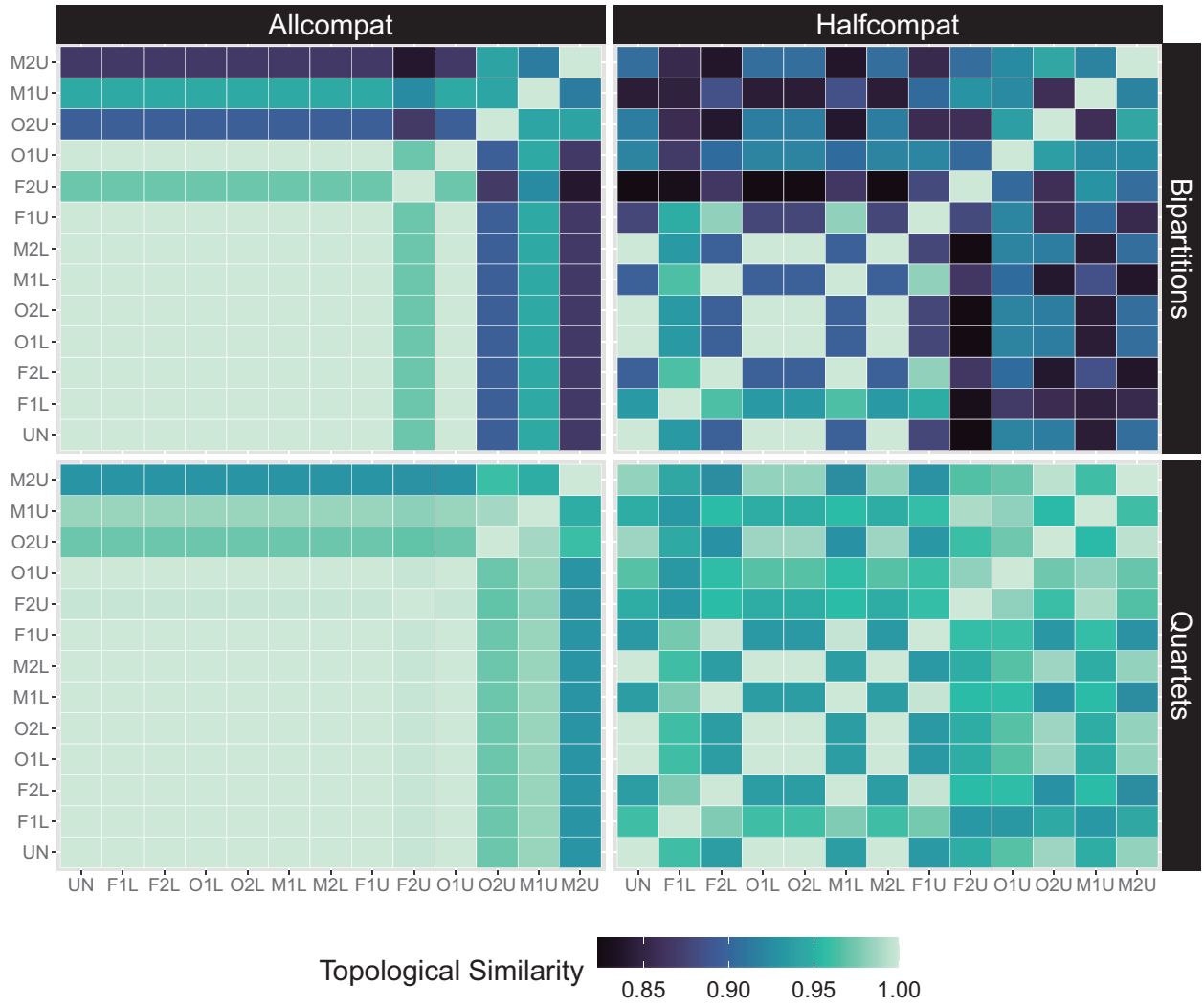

FIGURE 7. Topological similarity for maximum compatibility (allcompat) and 50% majority-rule consensus trees (halfcompat) obtained with the alternative models evaluated for the Cingulata dataset.

TABLE 2. Proportion of resolved bipartitions and quartets for the alternative models evaluated for the Cingulata dataset

Model	Bipartitions	Quartets
UN	0.67	0.83
F1L	0.75	0.91
F2L	0.79	0.95
O1L	0.67	0.83
O2L	0.67	0.83
M1L	0.79	0.95
M2L	0.67	0.83
F1U	0.75	0.95
F2U	0.83	0.91
O1U	0.71	0.88
O2U	0.67	0.84
M1U	0.75	0.90
M2U	0.75	0.85

Porto et al. 2021; Casali et al. 2022). Model selection is undeniably an important step of statistical phylogenetic reconstructions, and Bayes factor reliability as a criterion to select appropriate partitioning schemes is well established (Brown and Lemmon 2007; Rosa et al. 2019; Casali et al. 2022). Nonetheless, a systematic assessment of the impact of partitioning on estimates of parameters of major interest in phylogenetic analyses—such as topological precision and accuracy—was lacking for morphological data, despite being much better understood for molecular data (Brown and Lemmon 2007; Kainer and Lanfear 2015).

Based on the results of simulated and vertebrate empirical datasets, we found that topological precision and accuracy were not much affected by anatomical partitioning. Overall, our results agree with previous findings that reported minor variations in tree topology using anatomical partitioning of empirical datasets, despite relevant differences in estimates of marginal likelihoods among models (Tarasov and Génier 2015; Rosa et al. 2019; Casali et al. 2022). Previous empirical studies have shown that some anatomical partitions can be preferred to unpartitioned analysis using Bayes factor criterion (Tarasov and Génier 2015; Rosa et al. 2019; Varela et al. 2019). Those results may initially

appear to be at odds with what we observed for topologies, but we are dealing with different aspects of the performance of partitioned models, and it is quite possible for a model to better fit the data with no impact on the topology or to produce only a slightly different topology if compared to those obtained with an unpartitioned model.

We should be careful interpreting the results of analyses applying the model x7U for dataset B, frequently associated with statistically significant results for its increased precision and diminished accuracy. These analyses were also disproportionately affected by convergence issues, which is most likely associated with the fact that this scheme includes two very small partitions (<20 characters and <10% of the dataset), which may have posed difficulties in estimating the greater number of parameters present in unlinked analyses. Previous investigations observed that even in less parameterized models applying linked branch lengths, very imprecise estimates of rate multipliers are obtained for small anatomical partitions (Casali et al. 2022).

Studies of empirical datasets that separately analyzed anatomical partitions recovered topologies significantly different between those partitions, at least for part of the datasets analyzed (Mounce et al. 2016; Sansom and Wills 2017; Li et al. 2020). We reanalyzed some of those datasets and showed that when we use combined (i.e., total evidence, Kluge 1989) analysis, those topological differences are accommodated in the same or very similar consensus topologies irrespective of whether or not we use anatomical partitions with linked branch lengths. Moreover, it is relevant to know that different partitions, if analyzed separately, would produce different topologies, but it does not necessarily imply that this would affect the topologies of partitioned models in combined analyses. In combined analyses, characters from different partitions interact and can reveal a hidden phylogenetic signal (Gatesy et al. 1999; Mounce et al. 2016), influencing the resulting topology. Unlinked branch lengths resulted in somewhat distinct topologies for empirical datasets, mostly due to their greater resolution. However, given the results of the simulations, this increased precision may not be a positive outcome, given the tendency of these models to negatively impact accuracy.

The case study of Cingulata allowed us to test more sophisticated hypotheses of anatomical partitioning. Modules obtained in morphometric analyses can be associated with evolutionary character covariation, which could ultimately reflect genetic and epigenetic pleiotropy and its interaction with natural selection or drift (Melo et al. 2016; Zelditch and Goswami 2021). Ontogenetic hypotheses for data partitioning are also conceptually interesting because they can relate to developmental constraints and biases, including shared heterochronic patterns (Koyabu et al. 2011, 2014), which would render characters interdependent to some extent. Despite that, partitioning schemes based on those criteria did not lead to distinct allcompat topologies in most of our analyses, a result aligned with those we obtained with simulated data and the vertebrate datasets. However, unlinked models sometimes returned different allcompat topologies, and different halfcompat topologies were obtained even with linked branch lengths, given the different degrees of resolution produced by some models. These complex partitioning hypotheses require expertise in the group being studied, so they were restricted here to a single dataset and did not allowed us to take general conclusions. It would be critical, nonetheless, that future studies with other datasets investigate how widespread this pattern is when applying similar partitioning schemes.

Comparison with Alternative Morphological Partition Schemes

In light of the results obtained here, in Rosa et al. (2019) and Casali et al. (2022), we can speculate that the partitioning of morphological datasets using anatomical subsets may not be the best approach to deal with the patterns of rate heterogeneity present in these datasets. Our results also suggest that phenotypic modularity may not be associated with anatomical partitions as they have been defined in phylogenetic studies.

Partitioning by homoplasy has been shown to perform consistently better than anatomical partitioning (Rosa et al. 2019), and methods that partition datasets using other proxies of evolutionary rates (e.g., as implemented in PartitionFinder2, Lanfear et al. 2016) also outperformed partitioning by anatomy for some datasets, although less consistently than when homoplasy is applied to segregate characters across partitions (Rosa et al. 2019). Casali et al. (2022) also observed that homoplasy-based partitions outperformed anatomical partitioning, with the latter returning very similar topologies to those obtained with unpartitioned models. It is straightforward to understand why this is the case since the primary goal of partitioning is to collectively model characters that share similar evolutionary patterns and rates, with their parameters estimated separately from those of other subsets of characters. Anatomical partitions may not produce well-segregated subsets since they are usually composed of characters of variable nature, like variations in shape, size, proportions, organization, and presence or absence of disparate kinds of structures (Casali et al. 2022). On the other hand, the degrees of homoplasy can work as direct proxies for the evolutionary rates (Rosa et al. 2019). Morphological characters may also be better segregated among partitions using other algorithmic approaches that directly focus on character rate variation (e.g., Azevedo et al. 2022).

Notwithstanding, methods using homoplasy or other rate-based metrics are necessarily topology dependent, and their sensitivity to a specific topology and the optimality criterion used to obtain it remains poorly explored (but see Felsinger 2019; Casali et al. 2022). Studies exploring homoplasy partitioning calculated the homoplasy indices with implied-weights parsimony with the parameter governing the strength of weighting

(*k*) set to the default value (Rosa et al. 2019; Brazeau et al. 2020; Lucena and Almeida 2021; Matos-Maraví et al. 2021). However, it has been recently shown that since alternative values for this parameter (or using equal weights parsimony) may lead to different topologies, this may result in different allocations of characters across partitions, which, in turn, can lead to different summary topologies in Bayesian inference (Casali et al. 2022). Additional studies will be necessary to further evaluate the magnitude of the influence of this initial topology on rate and homoplasy-based data partitioning methods, including simulations and more comprehensive systematic evaluations of empirical datasets.

Comparison with Molecular Data Partitioning

The use of partitioned models in molecular phylogenetics is a well-established practice (Brandley et al. 2005; Blair and Murphy 2011; Kainer and Lanfear 2015), and different criteria have been applied to define those data subsets, like genes across the genome, introns × exons, codon positions, and stem × loop regions (Brandley et al. 2005; Kainer and Lanfear 2015). Conceptually, partitioning constitutes the same procedure for morphological and molecular data. In both cases, the objective of this approach is to separately infer parameters for data subsets that were hypothesized to evolve—to some extent—independently from the other subsets of a given dataset (Clarke and Middleton 2008; Lanfear et al. 2012). In practice, however, quite distinct substitution models are often selected and applied to alternative molecular partitions (Lemmon and Moriarty 2004), whereas for morphology, all partitions usually have characters modeled by the Mk model. Although an extension of this model allowing frequency asymmetry among states is available (Wright et al. 2016), it has been rarely used (e.g., Simões et al. 2020; May et al. 2021). The asymmetrical model is more prone to overparameterization and poses difficulties in achieving convergence, which may discourage its use (Simões et al. 2020). An alternative asymmetric Mk model (Pyron 2017) or a nonstationary Mk model (Klopfstein et al. 2015) have also been proposed but are seldom used. In that way, in practice, morphological data partitions differ in fewer and, probably, less consequential parameters, potentially explaining part of the disagreements between our results and those obtained by evaluating molecular data partitioning.

Differently from what we observed here for morphology, for molecular datasets, alternative partitioning usually leads to moderate to substantial impacts on tree topology (Nylander et al. 2004; Kainer and Lanfear 2015). In another study, also using molecular data, topological differences were more modest and mostly associated with weakly supported nodes (Brandley et al. 2005). An improvement in resolution, which is directly related to the increase in node support in Bayesian inference, was reported for partitioned analyses of molecular datasets (e.g., Brandley et al. 2005). In contrast, some other studies reported a general decrease in node support, but with greater differences observed only in clades presenting lower support values (e.g., Powell et al. 2013). Systematic assessments of node support for simulated molecular data using partitioned models and Bayesian inference indicates that mismodeling can increase the variance of estimated posterior probabilities, without a trend for higher or lower values (Brown and Lemmon 2007).

Here, on the other hand, we observed that mismodeling is not much consequential. Topological resolution is mainly affected when branches are unlinked during analyses, irrespective of how data were simulated or the number of partitions being considered in analyses. However, this improvement in precision is accompanied by a decrease in accuracy; and hence, it is an undesirable property of these models. Linked branch lengths have been better supported for most molecular (Duchêne et al. 2020) and morphological datasets (Tarasov and Génier 2015; Rosa et al. 2019), which may indicate that these less parametrized models may, in many instances, capture the shared temporal component of branch lengths common to all partitions and characters. This also suggests that the effects of heterotachy may be less important at the level of anatomical modules, even though they have been shown to be relevant at the level of individual morphological characters (but not so much for molecular data, Goloboff et al. 2018a).

We also observed that under- and overpartitioning have equally negligible impacts on topological precision and accuracy. This is quite different from the results observed for empirical molecular data, for which it was observed that the impacts of ignoring partitioning are more severe than using incorrectly underpartitioned schemes (Kainer and Lanfear 2015). Simulations showed a continuous loss of precision in estimates of node support from correct to gradually under or overpartitioned schemes for molecular data (Brown and Lemmon 2007), with a clear pattern of loss of precision when wrong partitioning schemes were applied.

Despite those incongruences observed between the performance of data partitioning for morphological and molecular datasets, it may be premature to assume that this can be fully explained as a difference in the nature of data used in the analyses. Other morphological partitioning criteria, particularly by homoplasy (Rosa et al. 2019), maybe more appropriate and achieve performance comparable to molecular partitioning, in the sense of affecting the inferred topologies more substantially, as shown by Casali et al. (2022).

Limitations of Simulations and Future Directions

Simulated data will always be limited in complexity if compared to empirical datasets. Despite that, simulation studies provide insights complementary to those obtained from the exploration of empirical data, and it is noteworthy that the general pattern which emerged here from simulations was also recovered for empirical

datasets, reinforcing our conclusions. However, the results obtained with the use of more complex anatomical partitions, as those applied to the Cingulata dataset, suggest that some partitioning hypotheses may be more consequential to topological inference than the typical anatomical modules usually applied in phylogenetic studies, even though this may also be a dataset-specific pattern; and hence, not generalizable.

The list of possible variables to be explored as potential interacting factors in a simulation is extensive and practical limitations should be considered, in order to focus on what seem to be the more relevant parameters for the matter at hand (Barido-Sottani et al. 2020). Here we adopted a method for data simulation directly informed by those parameters and properties of empirical datasets that seemed more relevant to simulate partitioned datasets. Nonetheless, the simulation of morphological data is a complex endeavor, and many important aspects were not considered directly here, such as the hierarchical relationship of characters (Tarasov 2019, 2020) or correlations and the presence of serial homology (Billet and Bardin 2019), to cite a few examples.

Adopting only two empirical datasets as references to simulate the data is another limitation of our study, although results were, in general, consistent between them. This is relevant, since they differ in many fundamental properties, like the number of taxa, number of partitions per partitioning scheme, partition sizes, and missing data distribution. Future simulation studies may benefit from exploring properties from more datasets and variations in those parameters, among others.

Lastly, we have focused on the influence of anatomical partitions on topologies of nonclock trees. There is a growing interest in co-estimating topologies and divergence times by applying tip-dating methods for morphological data, and clock partitions have been increasingly used in these analyses (Lee 2016; Zhang and Wang 2019; Simões et al. 2020; Simões and Pierce 2021). Simulation studies will be necessary to evaluate how partitions may influence not only topological precision and accuracy, but other parameters as well, like divergence times and evolutionary rates in these analyses and in analyses combining morphological and molecular data.

Conclusion

Our results indicate that partitioning by anatomy has, overall, a minor influence on summary topologies while conducting Bayesian phylogenetic analyses of morphological data and that models with unlinked branch lengths should be used with caution when applied alongside anatomical partitions, given their tendency to produce more resolved but less accurate consensus topologies. However, we should bear in mind that a few empirical datasets have been influenced by anatomical partitioning, and the performance of these models in the inference of other parameters, like partition-specific evolutionary rates was not explored here, so it would not be reasonable to entirely discourage their use. That being considered, it is likely that other ways of partitioning data may perform better than the anatomical criterion, such as using character's homoplasy indices or rates to define partitions. Researchers should also consider the costs—in time and resources—and the potential benefits of exploring alternative partition schemes in their empirical studies, and hopefully, this study can provide some guidance in this evaluation.

Supplementary Data

Data available from the Dryad Digital Repository: http://dx.doi.org/10.5061/dryad.vmcvdncv1

Funding

This work was supported by Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES-Finance Code 0001). Also, this work was partly supported by the grants #2018/09666-5 and #2022/00044-7, São Paulo Research Foundation (FAPESP) and by the Brazilian National Council for Scientific and Technological Development (CNPq grant 422019/2018-6).

Acknowledgments

We would like to thank Eduardo A. B. Almeida for making computational resources available for preliminary analyses of this study. We also would like to thank Cayo Dias, Brunno Rosa, Joseph Keating, Thomas Guillerme, Edward Susko, James Albert, Stephen Smith, Bryan Carstens, and three anonymous reviewers for suggestions that contributed significantly to improve the manuscript. We are also indebted to Daniel Barasoain for kindly making available the input nexus file for the Cingulata dataset. This research was developed with HPC resources made available by "Superintendência de Tecnologia da Informação" from "Universidade de São Paulo."

REFERENCES

Azevedo G.H.F., Bougie T., Carboni M., Hedin M., Ramírez M.J. 2022. Combining genomic, phenotypic and Sanger sequencing data to elucidate the phylogeny of the two-clawed spiders (Dionycha). Mol. Phylogenet. Evol. 166:107327.

Barasoain D., González Ruiz L., Tomassini R., Zurita A., Contreras V., Montalvo C. 2021. First phylogenetic analysis of the *Miocene armadillo Vetelia* reveals novel affinities with Tolypeutinae. Acta Palaeontol. Pol. 66:s031–s046.

Barido-Sottani J., Saupe E.E., Smiley T.M., Soul L.C., Wright A.M., Warnock R.C.M. 2020. Seven rules for simulations in paleobiology. Paleobiology 46:435–444.

Beer D., Beer A. 2019. phylotate: Phylogenies with annotations. https://cran.r-project.org/package=phylotate.
Billet G., Bardin J. 2019. Serial homology and correlated characters in morphological phylogenetics: modeling the evolution of dental crests in placentals. Syst. Biol. 68:267–280.
Blair C., Murphy R.W. 2011. Recent trends in molecular phylogenetic analysis: where to next? J. Hered. 102:130–138.
Bonferroni C. 1936. Teoria statistica delle classi e calcolo delle probabilita. Pubbl. del R Ist. Super. di Sci. Econ. e Commericiali di Firenze. 8:3–62.
Brandley M.C., Schmitz A., Reeder T.W. 2005. Partitioned Bayesian analyses, partition choice, and the phylogenetic relationships of scincid lizards. Syst. Biol. 54:373–390.
Brazeau M.D., Giles S., Dearden R.P., Jerve A., Ariunchimeg Y.A., Zorig E., Sansom R., Guillerme T., Castiello M. 2020. Endochondral bone in an Early Devonian 'placoderm' from Mongolia. Nat. Ecol. Evol. 4:1477–1484.
Brown J.M., Lemmon A.R. 2007. The importance of data partitioning and the utility of Bayes factors in Bayesian phylogenetics. Syst. Biol. 56:643–655.
Casali D.M., Boscaini A., Gaudin T.J., Perini F.A. 2022. Reassessing the phylogeny and divergence times of sloths (Mammalia: Pilosa: Folivora), exploring alternative morphological partitioning and dating models. Zool. J. Linn. Soc. 0:0–0.
Clarke J.A., Middleton K.M. 2008. Mosaicism, modules, and the evolution of birds: results from a Bayesian approach to the study of morphological evolution using discrete character data. Syst. Biol. 57:185–201.
Close R.A., Friedman M., Lloyd G.T., Benson R.B.J. 2015. Evidence for a mid-Jurassic adaptive radiation in mammals. Curr. Biol. 25:2137–2142.
Duchêne D.A., Tong K.J., Foster C.S.P., Duchêêne S., Lanfear R., Ho S.Y.W. 2020. Linking branch lengths across sets of loci provides the highest statistical support for phylogenetic inference. Mol. Biol. Evol. 37:1202–1210.
Felsinger S.M. 2019. Untangling the tree of life: Which partitioning strategies improve phylogenetic inference? [Masters Thesis]. Durham: Durham University.
Gatesy J., O'Grady P., Baker R.H. 1999. Corroboration among data sets in simultaneous analysis: hidden support for phylogenetic relationships among higher level artiodactyl taxa. Cladistics 15:271–313.
Goloboff P.A., Arias J.S. 2019. Likelihood approximations of implied weights parsimony can be selected over the Mk model by the Akaike information criterion. Cladistics 35:695–716.
Goloboff P.A., Pittman M., Pol D., Xu X. 2018a. Morphological data sets fit a common mechanism much more poorly than DNA sequences and call into question the Mkv model. Syst. Biol. 68:494–504.
Goloboff P.A., Torres A., Arias J.S. 2018b. Weighted parsimony outperforms other methods of phylogenetic inference under models appropriate for morphology. Cladistics 34:407–437.
Guillerme T. 2018. dispRity: a modular R package for measuring disparity. Methods Ecol. Evol. 9:1755–1763.
Harrison L.B., Larsson H.C.E. 2015. Among-character rate variation distributions in phylogenetic analysis of discrete morphological characters. Syst. Biol. 64:307–324.
Heibl C. 2008. PHYLOCH: R language tree plotting tools and interfaces to diverse phylogenetic software packages. http://www.christophheibl.de/Rpackages.html.
Hillis D.M. 1995. Approaches for assessing phylogenetic accuracy. Syst. Biol. 44:3–16.
Jukes T., Cantor C.R. 1969. Evolution of protein molecules. In: Munro H.., editor. Mammalian protein metabolism. New York: Academic Press. p. 21–132.
Kainer D., Lanfear R. 2015. The effects of partitioning on phylogenetic inference. Mol. Biol. Evol. 32:1611–1627.
Kardong K. V. 2012. Vertebrates: comparative anatomy, function, evolution. New York: McGraw-Hill.
Keating J.N., Sansom R.S., Sutton M.D., Knight C.G., Garwood R.J. 2020. Morphological phylogenetics evaluated using novel evolutionary simulations. Syst. Biol. 69:897–912.
Klopfstein S., Vilhelmsen L., Ronquist F. 2015. A nonstationary markov model detects directional evolution in hymenopteran morphology. Syst. Biol. 64:1089–1103.
Kluge A.G. 1989. A concern for evidence and a phylogenetic hypothesis of relationships among Epicrates (Boidae, Serpentes). Syst. Zool. 38:7.
Koyabu D., Endo H., Mitgutsch C., Suwa G., Catania K.C., Zollikofer C.P.E., Oda S., Koyasu K., Ando M., Sánchez-Villagra M.R. 2011. Heterochrony and developmental modularity of cranial osteogenesis in lipotyphlan mammals. Evodevo 2:21.
Koyabu D., Werneburg I., Morimoto N., Zollikofer C.P.E., Forasiepi A.M., Endo H., Kimura J., Ohdachi S.D., Truong Son N., Sánchez-Villagra M.R. 2014. Mammalian skull heterochrony reveals modular evolution and a link between cranial development and brain size. Nat. Commun. 5:3625.
Krmpotic C.M., Nishida F., Galliari F.C., Pombo M.T., Acuña F., Barbeito C.G., Carlini A.A. 2021. The dorsal integument of the southern long-nosed Armadillo *Dasypus hybridus* (Cingulata, Xenarthra), and a possible neural crest origin of the osteoderms. Discussing evolutive consequences for amniota. J. Mamm. Evol. 28:635–645.
Lanfear R., Calcott B., Ho S.Y.W., Guindon S. 2012. PartitionFinder: combined selection of partitioning schemes and substitution models for phylogenetic analyses. Mol. Biol. Evol. 29:1695–1701.
Lanfear R., Frandsen P.B., Wright A.M., Senfeld T., Calcott B. 2016. PartitionFinder 2: new methods for selecting partitioned models of evolution for molecular and morphological phylogenetic analyses. Mol. Biol. Evol. 34:msw260.
Lee M.S.Y. 2016. Multiple morphological clocks and total-evidence tip-dating in mammals. Biol. Lett. 12:20160033.
Lee M.S.Y., Palci A. 2015. Morphological phylogenetics in the genomic age. Curr. Biol. 25:R922–R929.
Lemmon A.R., Moriarty E.C. 2004. The importance of proper model assumption in Bayesian phylogenetics. Syst. Biol. 53:265–277.
Lewis P.O. 2001. A likelihood approach to estimating phylogeny from discrete morphological character data. Syst. Biol. 50:913–925.
Li Y., Ruta M., Wills M.A. 2020. Craniodental and postcranial characters of non-avian Dinosauria often imply different trees. Syst. Biol. 69:638–659.
Lucena D.A.A., Almeida E.A.B. 2021. Morphology and Bayesian tip-dating recover deep Cretaceous-age divergences among major chrysidid lineages (Hymenoptera: Chrysididae). Zool. J. Linn. Soc. 0:1–44.
Maier W., Ruf I., Ruf I. 2016. Evolution of the mammalian middle ear: a historical review. J. Anat. 228:270–283.
Marshall D.C., Simon C., Buckley T.R. 2006. Accurate branch length estimation in partitioned Bayesian analyses requires accommodation of among-partition rate variation and attention to branch length priors. Syst. Biol. 55:993–1003.
Matos-Maraví P., Wahlberg N., Freitas A.V.L., Devries P., Antonelli A., Penz C.M. 2021. Mesoamerica is a cradle and the Atlantic Forest is a museum of Neotropical butterfly diversity: insights from the evolution and biogeography of Brassolini (Lepidoptera: Nymphalidae). Biol. J. Linn. Soc. 133:704–724.
May M.R., Contreras D.L., Sundue M.A., Nagalingum N.S., Looy C.V., Rothfels C.J. 2021. Inferring the total-evidence timescale of marattialean fern evolution in the face of model sensitivity. Syst. Biol. 0:1–22.
Melo D., Porto A., Cheverud J.M., Marroig G. 2016. Modularity: genes, development, and evolution. Annu. Rev. Ecol. Evol. Syst. 47:463–486.
Mounce R.C.P., Sansom R., Wills M.A. 2016. Sampling diverse characters improves phylogenies: craniodental and postcranial characters of vertebrates often imply different trees. Evolution 70:666–686.
Noden D.M., Trainor P.A. 2005. Relations and interactions between cranial mesoderm and neural crest populations. J. Anat. 207:575–601.
Novacek M.J. 1993. Patterns of Diversity in the Mammalian Skull. In: Hanken J., K. H.B., editors. The skull. Chicago (IL): University of Chicago Press. p. 438–545.
Nylander J.A.A., Ronquist F., Huelsenbeck J.P., Nieves-Aldrey J.L. 2004. Bayesian phylogenetic analysis of combined data. Syst. Biol. 53:47–67.
O'Reilly J.E., Puttick M.N., Parry L., Tanner A.R., Tarver J.E., Fleming J., Pisani D., Donoghue P.C.J. 2016. Bayesian methods outperform parsimony but at the expense of precision in the

estimation of phylogeny from discrete morphological data. Biol. Lett. 12:20160081.
O'Reilly J.E., Puttick M.N., Pisani D., Donoghue P.C.J. 2018. Probabilistic methods surpass parsimony when assessing clade support in phylogenetic analyses of discrete morphological data. Palaeontology 61:105–118.
Plummer M., Best N., Cowles K., Vines K. 2006. CODA: convergence diagnosis and output analysis for MCMC. R News 6:7–11.
Porto A., de Oliveira F.B., Shirai L.T., De Conto V., Marroig G. 2009. The evolution of modularity in the mammalian skull I: morphological integration patterns and magnitudes. Evol. Biol. 36:118–135.
Porto D.S., Almeida E.A.B., Pennell M.W. 2021. Investigating morphological complexes using informational dissonance and Bayes factors: a case study in corbiculate bees. Syst. Biol. 70:295–306.
Powell A.F.L.A., Barker F.K., Lanyon S.M. 2013. Empirical evaluation of partitioning schemes for phylogenetic analyses of mitogenomic data: an avian case study. Mol. Phylogenet. Evol. 66:69–79.
Prevosti F.J., Chemisquy M.A. 2009. The impact of missing data on real morphological phylogenies: influence of the number and distribution of missing entries. Cladistics 26:326–339.
Puttick M.N., O'Reilly J.E., Oakley D., Tanner A.R., Fleming J.F., Clark J., Holloway L., Lozano-Fernandez J., Parry L.A., Tarver J.E., Pisani D., Donoghue P.C.J. 2017284. Parsimony and maximum-likelihood phylogenetic analyses of morphology do not generally integrate uncertainty in inferring evolutionary history: a response to Brown et al. Proc. R. Soc. B Biol. Sci. 284: 20171636.
Puttick M.N., O'Reilly J.E., Pisani D., Donoghue P.C.J. 2019. Probabilistic methods outperform parsimony in the phylogenetic analysis of data simulated without a probabilistic model. Palaeontology 62:1–17.
Pyron R.A. 2017. Novel approaches for phylogenetic inference from morphological data and total-evidence dating in squamate reptiles (lizards, snakes, and amphisbaenians). Syst. Biol. 66:38–56.
R Core Team. 2022. R: A language and environment for statistical computing. https://www.r-project.org/.
Ronquist F., Teslenko M., van der Mark P., Ayres D.L., Darling A., Höhna S., Larget B., Liu L., Suchard Marc A., Huelsenbeck John P., Suchard M., Huelsenbeck J.P. 2012. MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. Syst. Biol. 61:539–542.
Rosa B.B., Melo G.A.R., Barbeitos M.S. 2019. Homoplasy-based partitioning outperforms alternatives in Bayesian analysis of discrete morphological data. Syst. Biol. 68:657–671.
Sansom R.S., Wills M.A. 2017. Differences between hard and soft phylogenetic data. Proc. R. Soc. B Biol. Sci 284:20172150.
Schrago C.G., Aguiar B.O., Mello B. 2018. Comparative evaluation of maximum parsimony and Bayesian phylogenetic reconstruction using empirical morphological data. J. Evol. Biol. 31:1477–1484.
Simões T.R., Caldwell M.W., Pierce S.E. 2020. Sphenodontian phylogeny and the impact of model choice in Bayesian morphological clock estimates of divergence times and evolutionary rates. BMC Biol. 18:1–30.
Simões T.R., Pierce S.E. 2021. Sustained high rates of morphological evolution during the rise of tetrapods. Nat. Ecol. Evol. 5:1403–1414.
Smith M.R. 2019a. Bayesian and parsimony approaches reconstruct informative trees from simulated morphological datasets. Biol. Lett. 15:20180632.
Smith M.R. 2019b. Quartet: comparison of phylogenetic trees using quartet and split measures. doi:10.5281/zenodo.2536318.
Smith M.R. 2020b. Information theoretic generalized Robinson–Foulds metrics for comparing phylogenetic trees. Bioinformatics 36:5007–5013.
Smith M.R. 2020a. TreeDist: distances between phylogenetic trees. R package version 2.4.0. Compr. R Arch. Netw.
Smith M.R. 2022. Robust analysis of phylogenetic tree space. Syst. Biol. 0:1–51.
Tarasov S. 2019. Integration of anatomy ontologies and Evo-Devo using structured Markov models suggests a new framework for modeling discrete phenotypic traits. Syst. Biol. 68:698–716.
Tarasov S. 2020. The invariant nature of a morphological character and character state: insights from gene regulatory networks. Syst. Biol. 69:392–400.
Tarasov S., Génier F. 2015. Innovative Bayesian and parsimony phylogeny of dung beetles (Coleoptera, Scarabaeidae, Scarabaeinae) enhanced by ontology-based partitioning of morphological characters. PLoS One 10:e0116671.
Varela L., Tambusso P.S., McDonald H.G., Fariña R.A. 2019. Phylogeny, macroevolutionary trends and historical biogeography of sloths: insights from a Bayesian morphological clock analysis. Syst. Biol. 68:204–218.
Warren D.L., Geneva A.J., Lanfear R. 2017. RWTY (R We There Yet): an R package for examining convergence of Bayesian phylogenetic analyses. Mol. Biol. Evol. 34:1016–1020.
Wiens J.J. 2004. The role of morphological data in phylogeny reconstruction. Syst. Biol. 53:653–661.
Wilcoxon F. 1945. Individual comparisons by ranking methods. Biometr. Bull. 1:80.
Wright A.M. 2019. A systematist's guide to estimating Bayesian phylogenies from morphological data. Insect Syst. Divers 3:1–14.
Wright A.M., Hillis D.M. 2014. Bayesian analysis using a simple likelihood model outperforms parsimony for estimation of phylogeny from discrete morphological data. PLoS One 9:e109210.
Wright A.M., Lloyd G.T. 2020. Bayesian analyses in phylogenetic palaeontology: interpreting the posterior sample. Palaeontology 63:997–1006.
Wright A.M., Lloyd G.T., Hillis D.M. 2016. Modeling character change heterogeneity in phylogenetic analyses of morphology through the use of priors. Syst. Biol. 65:602–611.
Xie W., Lewis P.O., Fan Y., Kuo L., Chen M.-H. 2011. Improving marginal likelihood estimation for Bayesian phylogenetic model selection. Syst. Biol. 60:150–160.
Yang Z. 1996. Among-site rate variation and its impact on phylogenetic analyses. Trends Ecol. Evol. 11:367–372.
Zelditch M.L., Goswami A. 2021. What does modularity mean? Evol. Dev. 23:377–403.
Zhang C., Wang M. 2019. Bayesian tip dating reveals heterogeneous morphological clocks in Mesozoic birds. R. Soc. Open Sci. 6:182062.