

Phylogenetic Signal and Bias in Paleontology

ROBERT J. ASHER^{1,*} AND MARTIN R. SMITH²

¹Department of Zoology, University of Cambridge, Downing Street, Cambridge CB2 3EJ, UK; and

²Department of Earth Sciences, Durham University, Lower Mount Joy, Durham DH1 3LE, UK

*Correspondence to be sent to: Department of Zoology, University of Cambridge, Downing Street, Cambridge CB2 3EJ, UK;
E-mail: r.asher@zoo.cam.ac.uk.

Received 9 February 2021; reviews returned 24 August 2021; accepted 27 August 2021

Associate Editor: Liliانا Davalos

Abstract.—An unprecedented amount of evidence now illuminates the phylogeny of living mammals and birds on the Tree of Life. We use this tree to measure the phylogenetic value of data typically used in paleontology (bones and teeth) from six data sets derived from five published studies. We ask three interrelated questions: 1) Can these data adequately reconstruct known parts of the Tree of Life? 2) Is accuracy generally similar for studies using morphology, or do some morphological data sets perform better than others? 3) Does the loss of non-fossilizable data cause taxa to occur in misleadingly basal positions? Adding morphology to DNA data sets usually increases congruence of resulting topologies to the well-corroborated tree, but this varies among morphological data sets. Extant taxa with a high proportion of missing morphological characters can greatly reduce phylogenetic resolution when analyzed together with fossils. Attempts to ameliorate this by deleting extant taxa missing morphology are prone to decreased accuracy due to long-branch artifacts. We find no evidence that fossilization causes extinct taxa to incorrectly appear at or near topologically basal branches. Morphology comprises the evidence held in common by living taxa and fossils, and phylogenetic analysis of fossils greatly benefits from inclusion of molecular and morphological data sampled for living taxa, whatever methods are used for phylogeny estimation. [Concatenation; fossilization; morphology; parsimony; systematics; taphonomy; total-evidence.]

Paleontology is by definition a field in which some data are missing. Anything that becomes a fossil has to go through a paleontological filter. Fossils generally lack genetic and soft tissue data, and their hard parts are often fragmentary and incomplete (Briggs 1995; Smith 1998; Donoghue and Purnell 2009). There is no doubt that the paleontological filter can make it difficult to infer the phylogenetic affinities of fossils. In some cases, it may render definitive statements about affinities impossible. While it is trivial to say that, at some point, loss of data will have a negative effect on phylogenetic accuracy, it is more difficult to accurately generalize about where that point is, or what the overall impact of the paleontological filter on systematics has been. Sansom and Wills (2013, 1) stated that the non-random process of taphonomy (i.e., that some body tissues are more likely to decay than others) “systematically distorts phylogeny” by artifactually placing extinct species closer to the root of a given evolutionary tree than they actually are. This interpretation has been accepted by at least some authors (Foley et al. 2016; Springer et al. 2019) despite several examples to the contrary (e.g., Murdock et al. 2014; Nanglu et al. 2015; Pattinson et al. 2015; Sansom 2016; Klompaker et al. 2017; Flannery-Sutherland et al. 2019).

The paleontological literature abounds with phylogenetic interpretation, and indeed the field depends on it to understand past life on Earth. Paleontologists therefore continue to use fossilizable, morphological data out of necessity, and in some cases do so along with samples of living taxa with molecular data (e.g., Eernisse and Kluge 1993; Brochu 1997; Gatesy and O’Leary 2001; Lee 2001; Asher et al. 2003, 2005; Wiens

et al. 2010; Ronquist et al. 2012a; O’Leary et al. 2013; Beck and Lee 2014; Klopstein et al. 2015; Koch and Parry 2020). However, these studies are exceptions. Most phylogenetic reconstructions focusing on fossil groups sample few or no living taxa; when living taxa are present, they usually lack non-fossilizable (typically molecular) data. This applies just as much to high-profile papers in *Nature* and *Science* (e.g., Baron et al. 2017; Huttenlocker et al. 2018; Zhou et al. 2019; Krause et al. 2020) as it does to recent articles in (for example) the *Journal of Vertebrate Paleontology* and *Journal of Systematic Paleontology*. Most paleontologically informed biologists nonetheless agree that anatomical data recoverable from fossils, and molecular data from living taxa, are mutually informative in establishing the timing and branching patterns of the Tree of Life (Smith 1998; Brochu 2003; Jenner 2004; Reisz and Müller 2004; Asher et al. 2005, 2008, 2009; Wiens 2009; Benton et al. 2009, 2015; Lee and Palci 2015; Pyron 2015).

Here, we explore the extent to which fossilizable data may be systematically misleading and quantify phylogenetic congruence with well-corroborated branches of the phylogenetic Tree of Life. We test for bias using six data sets derived from five phylogenetic studies: four of mammals (Asher 2007; Pattinson et al. 2015; Huttenlocker et al. 2018; Halliday et al. 2019) and one of birds (Livezey and Zusi 2006), all of which we combine with alignments of DNA from corresponding taxa (Table 1). We simulate the process of fossilization for each data set, and compare resulting topologies with those derived from independent, genomic studies. This allows us to assess the contributions of morphology

and DNA to phylogenetic accuracy, and measure the detrimental influence of fossilization.

MATERIALS AND METHODS

Character Matrix Assembly

Each of our data sets required varying degrees of editing in order to combine with available molecular data and yield results pertinent to our research questions. We combined the morphological data from Asher (2007), Huttenlocker et al. (2018), and Halliday et al. (2019) with corresponding genera from the alignment of Upham et al. (2019), which consisted of 39,099 DNA sites (Supplementary Table S1 available on Dryad at <http://dx.doi.org/10.5061/dryad.w3r2280q3>). Morphological matrices for birds (Livezey and Zusi 2006; Supplementary Table S1 available on Dryad) and primates (Pattinson et al. 2015) lacked sufficient overlap with Upham et al. (2019); we therefore combined them with DNA sequence alignments from (respectively) Prum et al. (2015) and Springer et al. (2012; also used in Pattinson et al. 2015). We did not alter any of the alignments. Given the varying taxon samples in our data sets and our reliance on equally weighted parsimony (justified below), presence of gapped sites for all taxa would not affect choice of an optimal topology, although there are differences in the number of parsimony-informative DNA sites and morphological characters in each data set, as indicated in Table 1.

We noted character coding mistakes in three of these matrices: Pattinson et al. (2015), Huttenlocker et al. (2018), and Halliday et al. (2019). In addition, Mayr (2008) critiqued many of the coding decisions of Livezey and Zusi (2006). In order to treat all of these studies equitably, we did not make any corrections to their published matrices. However, the phylogenetic matrices provided by Halliday et al. (2019) had a particularly large number of anomalies, including large blocks of missing data for extant taxa (as detailed in Supplementary Appendix S1 available on Dryad). We therefore used a further matrix (“Halliday-50”) retaining only the 32 extant genera that were sampled for at least 50% of the characters in the “S1 Morphological Matrix” of Halliday et al. (2019), and that are also at least 50% complete in the alignment of Upham et al. (2019). Supplementary Appendix S1 available on Dryad provides further details on each morphological matrix with its respective DNA alignment; Supplementary Appendix S2 available on Dryad provides our taxon-character matrices in TNT format. Supplementary Table S1 available on Dryad provides information on taxon overlap of our morphological matrices with the alignments of Prum et al. (2015) and Upham et al. (2019).

The Well-Corroborated Tree and Data Independence

There is now a huge body of data supporting the basic phylogenetic topology for mammals and birds

(Fig. 1), including genomic alignments of primarily exons (Pattinson et al. 2015), introns and other non-coding regions (Reddy et al. 2017), microRNAs (Tarver et al. 2016), and ultraconserved elements (Esselstyn et al. 2017). Even given uncertainty at certain branches, for example the placement of tree shrews and bats among placental mammals or relationships among neoavian birds, an evolutionary tree can still be well-corroborated yet acknowledge uncertainty via polytomies. For our study, the basic topology for mammals derives from the UCE data set of Esselstyn et al. (2017) and the genomic and microRNA data set of Tarver et al. (2016). Their topology for primates is supported (with a better taxon sample) by the alignment of 69 nuclear and 10 mitochondrial genes by Springer et al. (2012). The topology for birds is based on a consensus of the Bayesian analysis of a 395 kb alignment of primarily coding exons by Prum et al. (2015: Fig. 1) with the optimal likelihood tree of 54 largely intronic loci for 235 bird species by Reddy et al. (2017: Fig. 3, their “early bird II” data set), which is in turn consistent with the topology for paleognaths figured by Yonezawa et al. (2017). There is no overlap between any of the morphological data sets and the data behind the well-corroborated trees shown in Figure 1.

Artificial Extinction (“ArtEx”)

To investigate the effect of taphonomic processes on phylogenetic inference, we applied artificial fossilization (Asher and Hofreiter 2006; Asher et al. 2008; Pattinson et al. 2015), or “ArtEx”. Each fossil taxon, or “template”, is missing all DNA and at least some hard-tissue characters. An extant subject taxon can be artificially “fossilized” by replacing each of its coded characters with “?” so as to correspond with missing characters in a given fossil template. Our methods to identify all “?” characters in a given fossil template, and change the corresponding characters to “?” in an extant subject, are similar to those of Pattinson et al. (2015), except for our use of R (R Core Team 2020) and TNT (Goloboff and Catalano 2016). Supplementary Table S2 available on Dryad provides further details on transforming extant taxa into artificial fossils, including R scripts. We acknowledge use of R packages ape (Paradis 2012), phangorn (Schliep 2011), Quartet (Sand et al. 2014; Smith 2019b), tidyverse (Wickham et al. 2019), TreeDist (Smith 2020), and TreeTools (Smith 2019a).

Each ArtEx TNT analysis included only extant taxa, one of which was artificially fossilized per run. For display purposes, we left the root taxon unchanged and did not include it among the subject (S)-template (T) combinations. Hence, for each of our data sets, we undertook (S-1)T phylogenetic analyses with TNT using parsimony as the optimality criterion, with all character changes equal.

TABLE 1. Summary of matrices used in this study

| Data set | Group of focus | Source | # taxa (extant, fossil) | #morph chars | Range fossil % complete | DNA align | % informative sites/chars |
|---|-------------------|--|-------------------------|--------------|-------------------------|--|---------------------------|
| Asher (2007) | Placental mammals | www.robertjasher.com/database/webcomb.nex | 50 (38, 12) | 196 | 53–92% | 39,099 sites from Upham et al. (2019) | 47/94 |
| Halliday et al. (2019) | Placental mammals | datadryad.org/stash/dataset/doi:10.5061/dryad.2kp7vr8 | 225 (51, 174) | 748 | 7–97% | 39,099 sites from Upham et al. (2019) | 52/99 |
| Halliday et al. (2019) only extant taxa >50% complete | Placental mammals | datadryad.org/stash/dataset/doi:10.5061/dryad.2kp7vr8 | 206 (32, 174) | 748 | 7–97% | 39,099 sites from Upham et al. (2019) | 30/99 |
| Huttenlocker et al. (2018) | Mesozoic mammals | Extracted from "Huttenlocker et al. Supplementary Materials and Methods" PDF | 117 (18, 99) | 538 | 2–86% | 394,684 sites from Upham et al. (2019) | 34/99 |
| Livezey and Zusi (2006) | Aves | CD included in Livezey and Zusi 2006 | 125 (86, 39) | 2954 | 4–76% | 394,684 sites from Prum et al. (2015) | 39/82 |
| Pattinson et al. (2015), Seiffert et al. (2009) | Primates | datadryad.org/stash/dataset/doi:10.5061/dryad.4g42q | 111 (26, 85) | 360 | 8–88% | 61,199 sites from Springer et al. (2012) | 21/98 |

Note: Columns for number of taxa and characters correspond to the values used in this study, not necessarily those from the original publications. Column “% informative sites/chars” shows percent informative DNA sites and morphological characters (respectively) and is based on only taxa with overlapping DNA and morphological characters. Matrices in TNT format are provided in [Supplementary Appendix S2](#) available on Dryad.

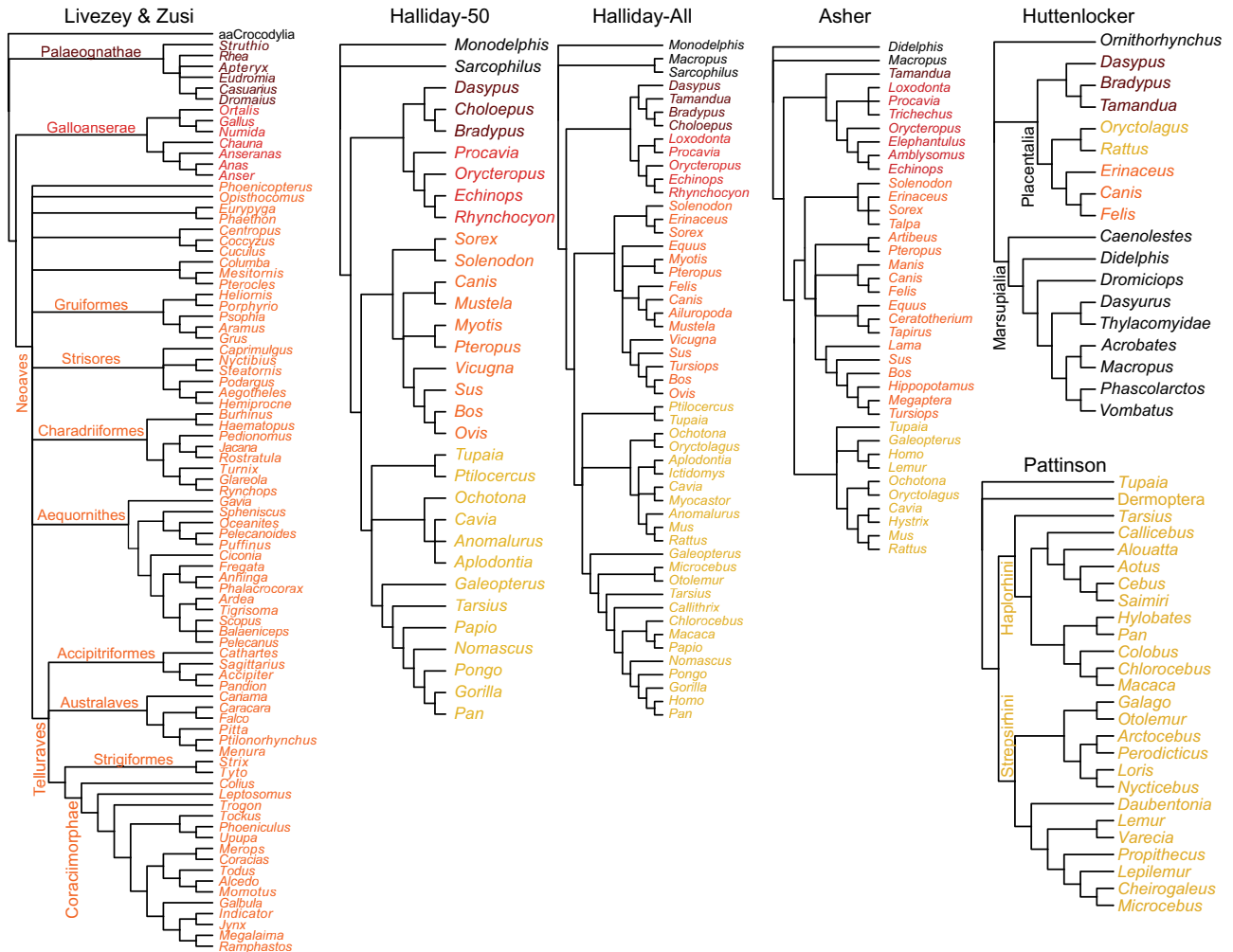


FIGURE 1. Well-corroborated trees. Morphological data sets are given at the top of each tree; taxa are pruned to match each morphological data set. Mammalian clades are shown in progressively lighter shades from Marsupialia and Monotremata (black) to Xenarthra, Afrotheria, Laurasiatheria, then Euarchontoglires; avian clades are rooted on Crocodylia (black) and progressively lighter shades from Paleognathae, Galloanserae, to Neoaves.

Character Randomizations

In order to assess the information content of real character states in each of our study matrices, we employed two strategies to remove phylogenetic information. One replaced real character states with binary 0/1 states chosen with equal probability (Supplementary Table S3A available on Dryad); the other replaced each real character state with a state drawn from a randomly selected taxon, using “TreeTools” (Smith 2019a). This procedure was repeated for every subject-template combination (excluding root taxa).

Subsampling DNA

By varying the amount of DNA included with morphological data in a given analysis, we sought to infer the dependence of a given data set on its DNA alignment (as opposed to morphology) to exhibit congruence with the well-corroborated tree. Toward this end, we

undertook two subsampling regimes, one of DNA sites (from 10% to 90% sites in 10% increments with 25 replicates each) and another of extant taxa known for DNA (from 20% to 80% taxa in 20% increments with 15 replicates each), using R scripts to write TNT batchfiles (Supplementary Table S3B, C available on Dryad). The number of replicates in sampling taxa was limited by the Huttenlocker data set, which had the smallest number of extant taxa (18 including the root taxon).

Calculating Congruence

The well-corroborated trees shown in Figure 1 comprised our reference topologies and accounted for the unique taxon sample of each data set. Analysis of many subject-template combinations yielded multiple most parsimonious trees. We calculated congruence values of most parsimonious topologies (MPTs) from each subject-template combination with the corresponding

well-corroborated tree in two ways: 1) using a strict consensus and 2) averaging across all bifurcating MPTs (Supplementary Table S4 available on Dryad). ArtEx topology congruence metrics excluded the root node which does not partition taxa into splits.

Compared to shared splits and Robinson–Foulds (RF) distance, quartet-based distance metrics are less sensitive to changes in the positions of a single taxon. Existing quartet measures (Estabrook et al. 1985; Day 1986; Smith 2019c) treat polytomies in a reference tree as a multifurcation event, and will penalize reconstructed trees that resolve such polytomies despite the fact that such resolution may be correct. In contrast, our new quartet similarity measure treats polytomies as expressions of uncertainty and corresponds to the average probability of guessing the true resolution of a quartet, based on its condition in a reconstructed tree. If a quartet is resolved in the well-corroborated tree, then its state in that tree is taken to be true with probability one; otherwise, any of the three possible resolutions of the quartet are considered equally likely to be true. The average score across all quartets is then given by

$$\text{raw score} = \frac{s + 1/3(r_r + r_{wc} + u)}{N_Q}$$

where s is the number of quartets resolved identically in the reconstructed and well-corroborated tree; r_r and r_{wc} are the number of quartets resolved only in the reconstructed or well-corroborated tree respectively; u is the number of quartets resolved in neither tree; and N_Q is the total number of quartet statements. We normalize this score such that the score of a tree that contains all of the quartet states of the well-corroborated tree is one, and the expected similarity of a randomly reconstructed tree, and the similarity of a completely unresolved tree, are both zero:

$$\text{maximum score} = \frac{s + d + r_{wc} + 1/3(r_r + u)}{N_Q}$$

$$\text{normalized score} = \frac{\text{raw score} - 1/3}{\text{maximum score} - 1/3}$$

We implemented this measure in the function “SimilarityToReference” in the R package “Quartet”. This measure rewards quartets that are correctly resolved in a reconstructed tree, without penalizing quartets that are resolved in the reconstructed tree but unresolved in the reference tree (i.e., excess resolution), and without treating the non-resolution of quartets in both trees as similarity. However, as with other quartet measures, it does not account for the non-independence of quartet statements.

Shared splits and quartet similarity are directly proportional with congruence; RF distance is inversely proportional. Given the varying sizes of our reference trees (Fig. 1), which range from 15 (Huttenlocker et al. 2018) to 70 (Livezey and Zusi 2006) total internal splits, we present shared splits as the proportion of

observed/potential shared with the respective reference tree, from 0 (none) to 1 (all).

Phylogenetic Search Strategies

Artificial extinction analyses alone required well over 40,000 distinct phylogenetic analyses. We also undertook phylogenetic analyses varying the proportion of DNA included per taxon (225 with and 225 without morphology for each of the six data sets), the number of taxa with DNA (60 for each data set), and analyzed full data sets and partitions therein. Each author (R.J.A. and M.R.S.) independently ran these phylogenetic analyses in TNT for each of our study matrices and interpretation and conclusions are based on results held in common. Newick trees provided in Supplementary Appendices S4 and S5 available on Dryad derive from the analyses run by R.J.A.

Despite its limitations (Felsenstein 1978; Smith 2019c), equal-weights parsimony serves as a proxy for more sophisticated tree reconstruction techniques, such as implied weighting or Bayesian methods, the use of which is impractical for the many thousands of phylogenetic analyses undertaken here. To be tractable in a limited time frame, each individual analysis has to finish within seconds or minutes, not hours or days. We found that a single phylogenetic analysis of even our smaller data sets (e.g., extant taxa in Halliday-50) still required many hours to approach convergence using a parallel version of MrBayes (v. 3.2.7, Ronquist et al. 2012b) on a computing cluster with appropriate models applied to codon positions, genes, and morphological partitions. Hence, we relied on equal-weights parsimony and searched treespace with a mix of strategies in TNT (Goloboff and Catalano 2016), as detailed in Supplementary Appendix S1 available on Dryad. Where available, we compared our results with optimal likelihood and Bayesian topologies described in the original publications of our test studies (Table 1). When needed, branch lengths were reconstructed with parsimony in PAUP* (Swofford 2003).

Measuring Phylogenetic Bias Resulting from Fossilization

Another goal of our study is to test the assertion that biological decay inherent in fossilization causes fossils to appear closer to the root than they really are (i.e., “stemward slippage” of Sansom and Wills 2013). With bifurcating ArtEx and reference trees, we used two approaches to quantify directional change of a given taxon before and after artificial fossilization. One followed Sansom and Wills (2013) and quantified the distance in nodes from the root to a given taxon; the other was the number of leaves (terminal taxa) in a given taxon’s sister group (function “SisterSize” in TreeTools). For each taxon, we defined its starting position as the average position in each of the MPTs returned under equal-weights parsimony including all data for extant taxa (Supplementary Appendix S3 available on Dryad).

Fully-bifurcating topologies are necessary because a consensus polytomy underestimates the number of nodes between the root and a given clade compared to a resolved topology. In five of six cases, parsimony applied to extant taxa from our morphology-DNA matrices yielded a single MPT; the Asher data set yielded three MPTs. We therefore used the fully-bifurcating majority-rule consensus to calculate starting positions for each extant taxon in the Asher data set. We defined the finishing position as each taxon's average position in each MPT recovered from every artificial extinction analysis in which that taxon was the subject. We averaged both sister group size and root-to-node distance first among all MPTs for a given subject-template combination, then among all templates for a given subject. This difference in starting and finishing position provided the basis to quantify directional change after deletion of non-fossilizable data, using both root-to-node distance and size of sister group.

Movement toward the root in a given topology is not necessarily synonymous with “stemward” movement, just as movement toward the most nested part in a given topology is not necessarily synonymous with “crownward” movement. Root-to-node distance can be influenced by changes in tree shape that are independent of the taxon of interest. For example, a taxon that “climbs” away from the root in one clade may still have a lower root-to-node distance if the overall position of the clade becomes less nested (Fig. 2). Moreover, the range of potential directional change is highly dependent on tree shape. The maximum root-to-node distance in a pectinate tree scales linearly with the number of leaves, but the maximum root-to-node distance in a balanced tree will be much smaller, increasing only with the logarithm of the number of leaves. This observation contradicts the assertion of Sansom and Wills (2013, 4) that “the null expectation for a perturbed single terminal taxon is an equal probability of displacement up or down a phylogeny, toward or away from the root in a fully resolved tree.” For example, in a balanced tree with 2^n leaves, all leaves exhibit maximum distance from the root; in such a tree, no taxon can move crownward as there is nowhere further from the root to go. A more appropriate null expectation is for the position of a taxon to become less predictable (more “random”) as data are removed, such that the removal of all data leaves a taxon equally likely to plot on any branch of the tree. To generate expected taxon placements under this “random null” model, we calculated the expected position of a randomly-placed taxon, calculated by averaging across all possible perturbations of removing the subject taxon from the tree, and computing the position measured if the subject taxon is added to each edge in this reduced tree in turn.

A separate possibility, related to the concept of long-branch error (e.g., Parks and Goldman 2014), is that incompletely coded taxa are most likely to plot where there is most uncertainty in the tree. An edge that is reconstructed as unambiguously exhibiting a single state has zero uncertainty; an edge on which any

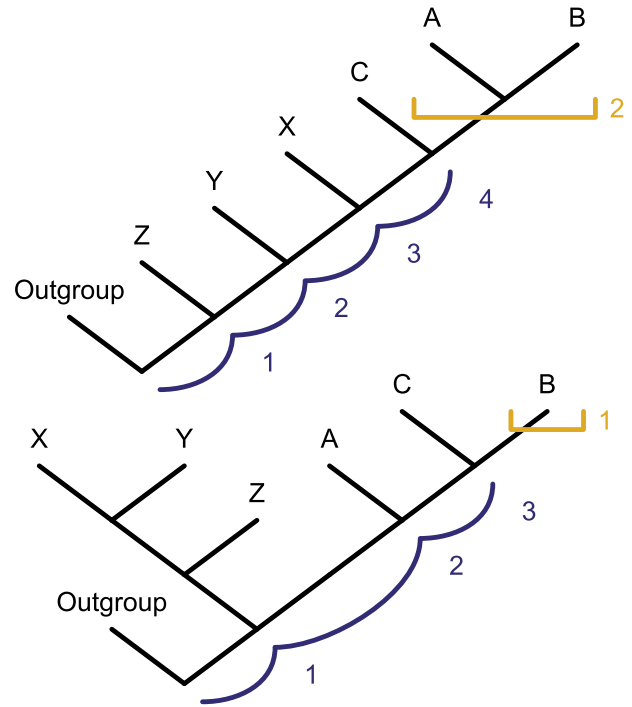


FIGURE 2. The range of possible values of the root-to-node distance is a function of the shape of the tree. In the top panel, leaf C falls in the stem group of clade (A, B) and has a large root-to-node distance (4). In the bottom panel, leaf C falls in the crown group of clade (A, B), yet has a smaller root node distance (3). Sister-group size (light shade) encapsulates the stemwardness/crownwardness of the subject leaf regardless of topological changes elsewhere in the tree.

available state could be parsimoniously reconstructed has maximum uncertainty. To construct an “uncertainty null” model, we quantified the uncertainty associated with each edge of the tree after removing a given subject by calculating the Shannon entropy of the set of character state symbols (i.e., “0”, “1”, etc.) that would not increase the parsimony score tree if they were exhibited by a taxon inserted at that edge. The entropy of a set of these symbols is given by $-\sum p \log(p)$, where p is the probability of observing each symbol. As symbols corresponding to character states that are observed in a majority of taxa are more likely to be observed, we obtained a value for p by dividing the number of taxa, excluding the subject, that bore each permitted symbol, by the number of taxa bearing any permitted symbol. We then used the uncertainty value associated with each edge to calculate a weighted mean stemwardness, by weighting the stemwardness value obtained by re-inserting a taxon at each edge in line with the uncertainty associated with that edge. Our quantification of sister-group size thus uses these two null expectations, “random” and “uncertainty,” to put directional bias after artificial fossilization in context.

We also quantified root-to-node distances to investigate directional bias. Due to the nature of our artificial extinction technique, which deletes non-fossilizable data for one taxon at a time, we expect (and have observed empirically) that topological changes resulting from

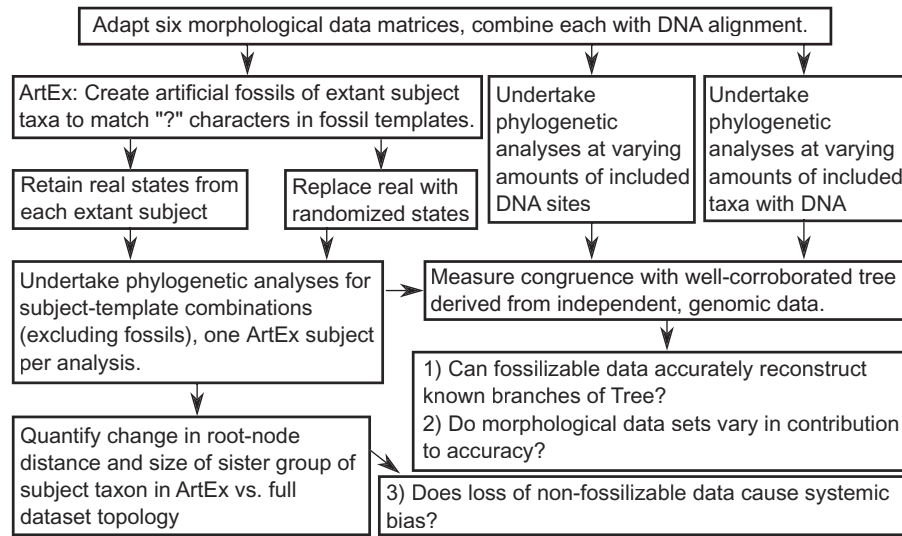


FIGURE 3. Flowchart illustrating the major elements of our methods and their relation to testing our main hypotheses.

this deletion will primarily influence each artificially fossilized taxon. Because other taxa often (but not necessarily) remain in the same starting positions, the shape of the initial, full-data tree is relevant to inform our expectations of how root-to-node distance could change following deletion of non-fossilizable data. A taxon that starts in a deeply nested position will have more potential to reduce its root-to-node distance after deletion of non-fossilizable data (i.e., “stemward slippage”) than a taxon starting in a basal position. Given the bifurcating topology from each full data set (Supplementary Appendix S3 available on Dryad), we therefore considered each taxon’s potential to directionally change along with our observed changes in root-to-node distance for that taxon.

Figure 3 provides a flowchart to summarize our methods and their relation to the hypotheses we test. Supplementary Table S5 available on Dryad gives our R script for calculating root-to-node distances. Supplementary Appendix S3 available on Dryad provides newick formatted representations of our well-corroborated trees and full data set topologies. Supplementary Appendix S4 available on Dryad provides MPTs resulting from ArtEx analyses, and Supplementary Appendix S5 available on Dryad strict consensus topologies of MPTs derived from each ArtEx subject-template combination.

RESULTS

Do Fossilizable Data Accurately Reconstruct Known Parts of the Tree of Life?

All of our data sets exhibited a significant correlation between quartet similarity and shared splits; both are

significantly correlated with increasing completeness of fossil templates (see the section on Artificial Extinction, below). In four of the six data sets (Asher, Huttenlocker, Pattinson, Halliday-50) and for all but the most incomplete templates, real and binary randomized character states fell in distinct, linear clusters on or near the same best fit line (Fig. 4) and real character states resulted in more congruence with the well-corroborated tree, as indicated by higher shared splits, higher quartet similarity, and lower RF, compared to randomized character states. The exceptions were the data sets of Halliday-All and Livezey-Zusi when we measured congruence with the well-corroborated tree using a strict consensus.

In Livezey-Zusi analyses with real character states, quartet similarity increased proportionally with shared splits but not RF distance (Fig. 4d). This was due to the wide range of RF distances among the least complete templates, many of which were poorly resolved with a corresponding lack of conflict represented by the RF distance metric. This behavior of RF distance was exacerbated by the unresolved neoavian comb in the well-corroborated tree for Aves (Fig. 1). Any increased resolution of Neoaves from more resolved templates would slightly increase RF distance to that otherwise well-corroborated tree, whether or not such resolution was correct. Another noteworthy aspect of the Livezey-Zusi data set is the extremely narrow range of similarity metrics for analyses using binary randomized character states for a given subject. As discussed in more detail below, nearly all of the Livezey-Zusi ArtEx results with binary randomized states were drawn to the longest branch, Crocodylia, which happens to occupy the root.

For the Halliday-All data set, strict-consensus topologies from analyses using randomized and real states overlapped considerably in terms of their congruence

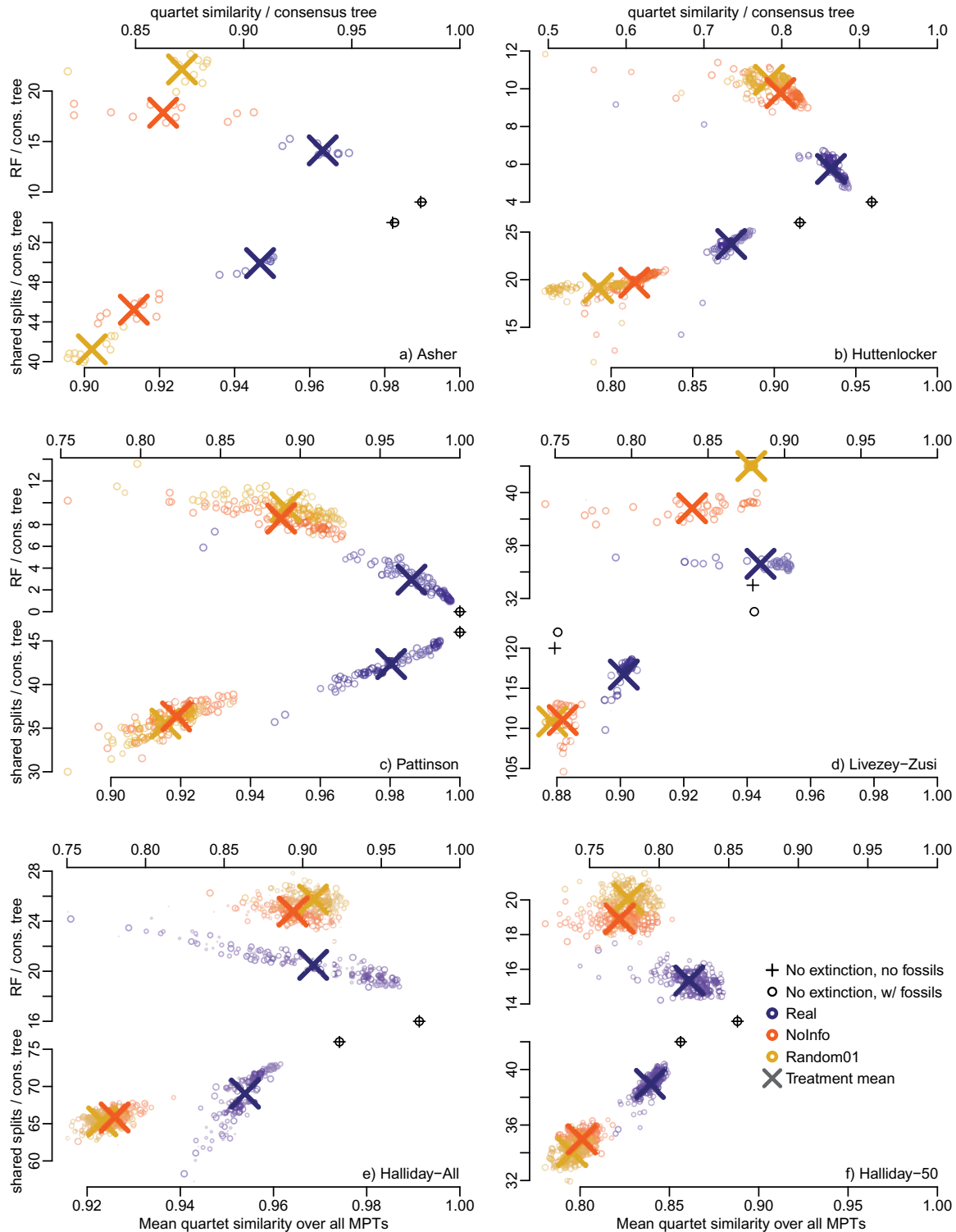


FIGURE 4. Correlation of congruence indices (Robinson–Foulds and proportion shared splits, Y-axes; quartet similarity, X-axes) representing the similarity of each data set’s ArtEx topologies to its respective well-corroborated tree. Each point depicts the average difference of topologies reconstructed after modifying the character scores for one subject taxon, averaged over all templates. In all cases, characters coded as “?” in a template taxon were coded as “?” in subjects. The remaining characters were either real states observed in extant subjects (dark shaded dots), replaced with a character state drawn from different extant taxa (medium shaded “NoInfo” dots), or randomly replaced with an equally probable “0” or “1” (light shaded “Random01” dots). Point size denotes the proportion of non-ambiguous states in the subject taxon, prior to treatment. Black crosses and circles indicate congruence values derived without artificial extinction from data sets of extant taxa only, and extant + fossil taxa, respectively.

with the well-corroborated tree (Fig. 4e). This contrasts with all of the other data sets, including Halliday-50, in which real versus randomized character states resulted in clusters with little or no overlap. This peculiar behavior of Halliday-All was due to three factors: 1) the high frequency of polytomies in strict consensus topologies and the resulting zero congruence from certain Halliday-All subject-template combinations with real states, 2) the masking of distinct areas of optimal tree space when measuring congruence with a strict consensus, as opposed to averaging a congruence index across all MPTs, and 3) the fact that average quartet similarity per template did not drop as much as average shared splits when including these instances of zero congruence. When averaging quartet similarity values across all MPTs for a given subject-template combination (and not calculating congruence based on a strict consensus thereof), all data sets exhibited higher quartet similarity values in all but the least complete templates, reflecting greater congruence with the well-corroborated tree using real versus random states (Fig. 4, lower X-axis).

Halliday-All (and no other data set) contained completely unresolved strict consensus topologies (with therefore zero congruence) because some of their extant subjects lacked key characters known for certain fossil templates. Even a fairly complete template such as *Ocepeia* (51% complete) combined with an incompletely coded extant taxon (*Macaca*, 33% complete) still resulted in a completely unresolved strict consensus. Other templates of similar overall completeness (e.g., *Pachyaena* at 50% complete) retained more characters that were also present in subjects like *Macaca*, and therefore returned more resolved topologies across all subjects. All subject-template combinations that returned congruence values (both shared splits and quartets) of “0” were from Halliday-All, and nearly two-thirds of these used real states. In some cases, subjects (*Homo*, *Oryctolagus*, *Tamandua*) returned higher quartet similarity values (but not shared splits or RF) using binary randomized states than real states, largely because quartet similarity does not change as much as shared splits or RF when one or few taxa are misplaced. The substantial overlap of quartet similarity scores resulting from strict-consensus representations of real versus randomized states in Halliday-All largely disappeared when we excluded their 19 morphologically least-coded (over 50% missing and inapplicable) extant subjects, as shown in our “Halliday-50” data set (Fig. 4f; see also [Guillerme and Cooper 2016a,b](#)).

Phylogenetic analyses of morphology alone.—Perhaps the simplest measure of phylogenetic signal in fossilizable, morphological data are the extent to which they recover well-corroborated nodes on their own. We applied equal weights and a series of homoplasy-based step weights (i.e., concavity values as implemented in TNT; [Goloboff et al. 2008](#)) to morphology-only data sets for each of the studies in Table 1, and measured congruence of strict

consensus topologies using shared well-corroborated splits and quartet similarity. In order from best to worst proportion of splits shared with the well-corroborated tree, analyses using only morphological data (and including fossils) were Pattinson, Huttenlocker, Livezey-Zusi, Halliday-50, Asher, and Halliday-All (Fig. 5a). Measured by quartet similarity, the relative performance of four of the six data sets changes. Pattinson and Huttenlocker including fossils are still among the best, but Livezey-Zusi and Halliday-All improve at most levels of implied weighting. Asher and Halliday-50 decrease in congruence across the board (Fig. 5b).

Figure 6 shows topologies based on an optimal implied weighting value applied to morphological data from Asher and Halliday-All, with similarities with their respective well-corroborated trees highlighted. Topologies derived from morphological characters in the Asher data set exhibit demonstrably incorrect placements for several taxa, usually consistent with homoplasy due to shared functional morphology (e.g., aquatic *Trichechus* with cetaceans) and with precedent in the older, comparative-anatomical literature (e.g., “Insectivora” including tenrecoids, “Volantia” including colugos and bats, “ungulates” including afrotheres and euungulates, “Edentata” including myrmecophagids and pangolins). The Halliday-All morphology data set recovers a smaller proportion of well-corroborated splits (0.18 vs. 0.3), all of which are either at or close to the tips (Fig. 6a). In Halliday-All, extant taxa with many missing characters appear in artificial “clades” with no historical precedent, such as *Macropus-Loxodonta* as sister to Glires, *Chlorocebus* outside of primates, and *Otolemur* with *Ailuropoda*. Inaccurate placement of these taxa substantially decreases congruence as measured by shared splits, and without these worst-coded extant taxa, morphology from Halliday-50 performs better (Fig. 5a). However, quartet similarity still rewards the Halliday-All data set for placing at least some members of well-corroborated clades in close proximity, such as the remaining primates, some carnivorans, euungulates (including cetaceans), and xenarthrans. Thus, in contrast to shared splits (Fig. 5a), where morphology from Halliday-All generates the least congruent topologies with no overlap with any of the others, it overlaps and sometimes exceeds morphological data sets of Asher and Halliday-50 when measured by quartet similarity (Fig. 5b).

Morphology combined with subsampled DNA.—Even when one type of data (such as fossilizable morphological characters) recovers relatively few well-corroborated groups by itself, it can still improve phylogenetic resolution and accuracy when combined with other data types ([Gatesy et al. 1999](#); [Asher et al. 2005](#); [Lee and Camens 2009](#); [Wiens 2009](#); [Thompson et al. 2012](#)), such as DNA. As described in Materials and Methods, we investigated the extent of this improvement with two subsampling regimes: increasing samples of DNA sites across all taxa and increasing proportions of taxa for all DNA sites.

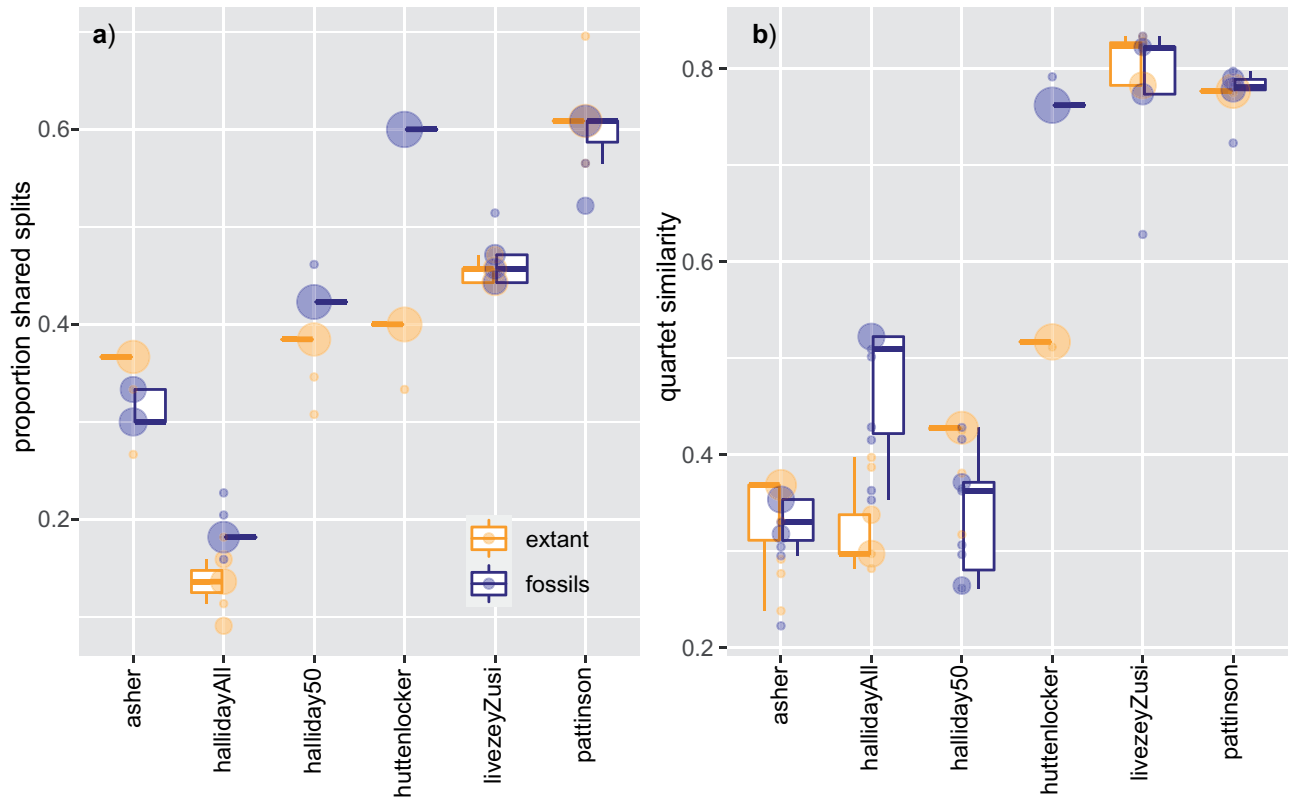


FIGURE 5. Proportion of shared splits (a) and quartet similarity (b) of topologies generated from equal and implied weighting using only morphological data with each data set's well-corroborated tree (Fig. 1). Concavity k -values ranged from 2, 4, 8, 16, 32, ... 999 and equal weighting, with size of circle proportional to number of analyses with the given Y-axis level of congruence. Most analyses generated a single MPT; in cases of multiple MPTs congruence was calculated using a strict consensus. Light shading indicates analyses with only extant taxa, dark shading extant taxa plus fossils. Horizontal lines show median and, where present, $1.5\times$ interquartile range (lines) and middle two quartiles (boxes).

All data sets exhibited at least a slight increase in shared splits with the well-corroborated tree using morphology plus increasingly large, random samples of DNA sites (Fig. 7, upper Y-axis); this was also observed without morphology for all data sets except birds. The effect on quartet similarity values of sequentially adding more DNA from each alignment was less pronounced but still slightly positive except in Halliday-50 and Livezey-Zusi (Fig. 7, lower Y-axis). The DNA alignment we combined with Livezey-Zusi (Prum et al. 2015) is by far the largest among our data sets; a 10% sample from its 395 kb is comparable to the entire 39 kb alignment of Upham et al. (2019), used to combined with the Asher, Huttenlocker and Halliday data sets. Perhaps relatedly, parsimony applied to 10–90% subsamples of the Prum et al. (2015) alignment, without morphology, showed nearly the same similarity to the well-corroborated tree in every case (Fig. 7). Hence, this data set did not show improvement with steadily increasing amounts of subsampled DNA alone and reached its asymptote of maximum congruence with just 20% of the 395 kb alignment.

The effect of adding morphology to subsampled DNA varied across data sets. In Asher and Pattinson, morphology increased similarity to the well-corroborated tree at nearly all levels of DNA sampling; in Halliday-All and

Huttenlocker, morphology had little effect; in Halliday-50 and Livezey-Zusi, the addition of morphological data mostly decreased the proportion of shared splits, but mostly increased mean quartet similarity without a positively trending slope as DNA subsamples increased in size (Fig. 7).

A reason for the downward trend of Halliday-50 (with 32 rather than 51 extant subjects sampled for DNA) when measured by quartet similarity, and for its lower levels of congruence compared to Halliday-All, was its tendency to root placentals incorrectly using parsimony, for example, with a paraphyletic Laurasiatheria represented by *Sorex* and *Solenodon*. The DNA alignment accompanying the Halliday-All data set more frequently placed the placental root within Afrotheria, which was usually accompanied by monophyly of Laurasiatheria, Euarchontoglires, and several clades therein. While this too resulted in paraphyly of some well-corroborated nodes (e.g., Afrotheria), overall it resulted in higher levels of congruence compared to Halliday-50 (Fig. 7). In addition, Halliday-50 had substantially fewer parsimony-informative sites (30% of the 39 kb alignment) compared to Halliday-All (52%), given the smaller taxon sample in Halliday-50 drawn from the DNA alignment of Upham et al. (2019, see Table 1).

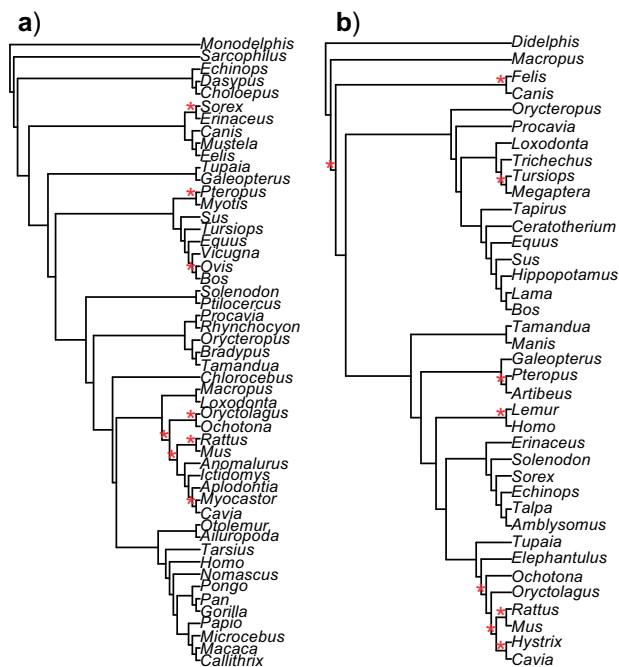


FIGURE 6. Optimal MPTs derived from morphological characters from Halliday-All (a) and Asher (b) data sets with implied weighting (concavity constant $k=64$) that maximize quartet similarity with their respective, well-corroborated trees (Fig. 1). Topologies were reconstructed using extant taxa and fossils, with fossils pruned for display purposes. Stars indicate splits held in common with the respective well-corroborated tree (8 of 44 in Halliday-All, 9 of 30 in Asher). Halliday-All shows higher quartet similarity (0.52 vs. 0.35) due to some near-misses (e.g., most but not all artiodactyls, primates, carnivorans).

None of our study sets contained exactly the same taxa, and two of them (Livezey and Zusi and Pattinson) used larger DNA alignments (395kb and 61kb, respectively) than the 39kb alignment from Upham et al. (2019) combined with the Asher, Halliday, and Huttenlocker morphology data sets (Table 1). Variation in the overall proportion of recovered splits is in most cases more indicative of the phylogenetic contributions of DNA than morphology; the influence of the latter is evident by comparing results with and without morphology added to each level of DNA sampling shown in Figure 7.

The addition of taxa known for DNA is positively correlated with both proportion of well-corroborated splits and quartet similarity, but the relationship is not linear (illustrated by the red diagonal in Fig. 8). Only Pattinson completely matches the well-corroborated topology when including all taxa known for DNA; the other data sets are missing at least some splits at 100% sampling. Below this, congruence measured by quartet similarity is in most cases disproportionately higher than the percent of taxa sampled for DNA; that is, nearly all levels of sampling taxa with DNA appear above the diagonal, except for Asher which drops below the line at 40% when calculating quartet similarity with strict consensus. Measured with shared splits, congruence is disrupted more by topological changes of individual taxa and is therefore lower in all data sets. Even so, when

measured by shared splits Pattinson is consistently above the diagonal; Halliday-All is consistently below it, and the others move from above to below the diagonal at either 40% (Asher) or 60% (Halliday-50, Huttenlocker, Livezey-Zusi).

Artificial Extinction (ArtEx).—All of our data sets show at least some improvement in congruence as fossil templates increase in completeness, measured by either shared splits or quartet similarity (Fig. 9). Except for Halliday-All (below), and using either a strict consensus or averaging across all MPTs, real character states perform better than either binary or no-info randomized character states. The Pattinson data set performs best, approaching complete congruence with the well-corroborated tree steadily throughout its range of 8–88% complete fossil templates (Fig. 9). In nearly all cases, mean congruence indices derived from real character states are prominently higher than means derived from randomized character states. As fossil completeness tends toward zero, the difference between real versus randomized treatments decreases, but the extent of this approach varies among our data sets.

Using strict consensus summaries, Halliday-All is the only data set to show extensive overlap of mean congruence values generated by real character states with those from randomized character states. Indeed, for many templates below 30% complete, mean congruence values based on real character states drop below those based on randomized states (Fig. 9e, top panel). As noted above, this is due to the uniquely high frequency of unresolved consensus topologies with correspondingly zero congruence. No such instances of zero congruence occur in the other data sets using either strict consensus or all-MPT averages.

Figure 10 and Table 2 summarize the data in Figures 4 and 9, allowing a comparison of the mean congruence of all subject-template combinations under each treatment. The difference in mean congruence between the artificial extinction and randomized treatments indicates the extent to which each morphological data set actively facilitates reconstruction of its associated well-corroborated tree. Livezey-Zusi showed the smallest difference, followed by Halliday-All, Halliday-50, Asher, Pattinson, and Huttenlocker with the largest (Table 2). In five of the six data sets, overlap in these ranges is due to low congruence of MPTs derived from real states for templates under 10% complete; Halliday-All shows more overlap in congruence of topologies generated by real versus random character states (Fig. 9). With the Pattinson, Huttenlocker, Asher and Halliday-50 data sets, real character states result in greater congruence with the well-corroborated tree under any similarity measure; this is less obviously the case in the Livezey-Zusi and Halliday-All data sets (Fig. 10; Table 2).

Overlap of mean congruence values derived from real versus randomized states in Livezey-Zusi partly reflects the particularly incomplete nature of their templates; only five of 39 templates were over 50%

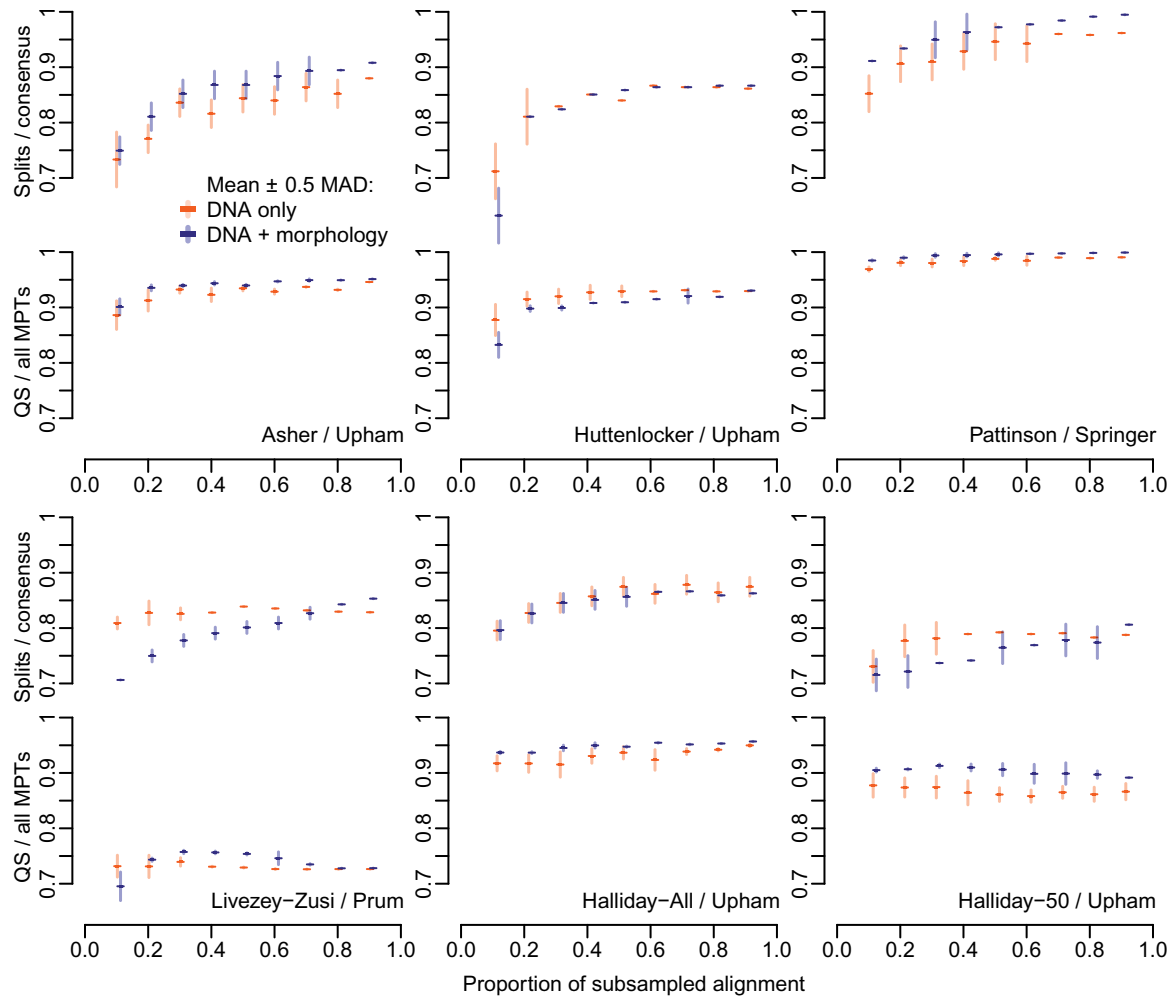


FIGURE 7. Proportion of well-corroborated splits using strict consensus (top Y-axes) and quartet similarity score averaged across all MPTs (bottom Y-axes) derived from 10% to 90% randomly sampled sites from DNA alignments (x-axis) with (dark shading) and without (light shading) morphology, applying jitter in x-direction to allow distinction between treatments. Horizontal lines represent means of 25 subsamples at each sampling level; vertical lines correspond to one median absolute deviation (0.5 above, 0.5 below) around each mean.

complete (Supplementary Fig. S1 available on Dryad). On the other extreme was Asher with all templates over 53% complete. Restricting the comparisons shown in Figure 10 to templates over 53% complete (Supplementary Fig. S2 available on Dryad), the differences in proportion shared partitions obtained by real versus 01-randomized character states increased, but the order of these differences was the same. Livezey-Zusi showed the smallest difference, followed by Halliday-All, Halliday-50, Asher, Pattinson, and Huttenlocker. When measured by quartets on either strict consensus or averaging across all MPTs, two switches in order occurred: between Asher and Halliday-50, and between Pattinson and Huttenlocker. Again, the overall order remained intact, with the latter two showing the largest difference in congruence obtained with real versus randomized character states, and Halliday-All and Livezey-Zusi the smallest (Table 2).

Are Fossils Reconstructed in Misleadingly Basal Positions?

Using either Sansom and Wills' (2013) measure of root-to-node distance or sister group size to quantify directional bias, our results do not support the claim that "fossilization causes organisms to appear erroneously primitive." Such a result can occur under certain circumstances, but this is neither a pervasive nor ubiquitous effect of fossilization itself.

Root-to-node distances.—When phylogenetically misplaced, artificially fossilized taxa often moved toward species sharing certain aspects of function or ecology. For example, the pangolin *Manis* and the xenarthran *Tamandua* possess a number of anatomical features resulting from their shared diet of ants and termites. In the combined DNA-morphology analyses of the Asher data set (Fig. 1), *Tamandua* was reconstructed close to

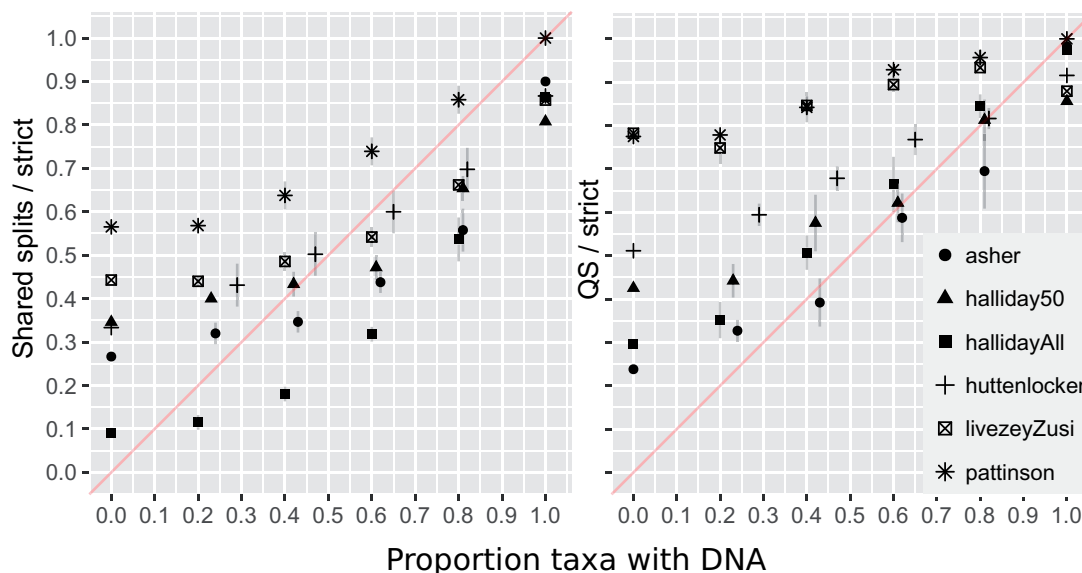


FIGURE 8. Congruence with the well-corroborated tree (Y-axis) as indicated by proportion of shared splits (left) and quartet similarity (right) recovered from strict consensus topologies using morphological data plus 0–100% samples of taxa known for DNA (X axis). Analyses at 0.0 and 1.0 on X-axis represent (respectively) single phylogenetic analyses of morphology-only and all DNA-morphology per data set. Samples in between are derived from 15 replicates with unique combinations of taxa known for DNA. Vertical lines represent one median absolute deviation (0.5 above, 0.5 below) around each mean; the diagonal represents a hypothetical, linear correlation of congruence with an increasing number of taxa sampled for DNA.

the root, whereas *Manis* appeared in a more nested position as sister to Carnivora. With its DNA deleted and using only *Tamandua* character states from fossil template characters in the Asher data set, *Tamandua* always appeared in a single MPT adjacent to *Manis*, within laurasiatheres, moving from two to 11 nodes from the root. Treating *Tamandua* as an artificial fossil therefore changed root-to-node distance by +9 (Fig. 11f). *Tamandua* exhibited the single greatest change in root-to-node distance among the 37 subject taxa in Asher (2007), followed by *Mus* with just under –6. Using the definitions of Sansom and Wills (2013), *Tamandua* exhibited “crownward” and *Mus* “stemward” slippage.

Given the shape of the full data topology for the Asher data set, *Tamandua* had more ways in which it could increase than decrease its root-to-node distance, as shown by the shaded region in Figure 11f. Assuming no changes in other taxa in the starting tree, the midpoint of its potential for change in the starting tree was just over 4 nodes from the root, so its observed change (+9) was indeed farther from the root given its available range, reflecting its attraction to a morphologically convergent species (*Manis*) located in a relatively nested part of the tree. Conversely, when treated as an artificial fossil, *Manis* always drifted toward *Tamandua*. Deletion of non-fossilized characters in *Manis* resulted in a reduction of root-to-node distance by –5, given its starting point as sister to Carnivora. With this initial tree shape, and assuming other taxa remain static, *Manis* could have shown an increase in root-to-node distance up to +4 and decrease of –7 nodes. Its observed change of –5 is thus closer to the root than the midpoint of its available range (–1.5).

The median of all observed stem-root changes added to the midpoints of their available areas of change for taxa in the Asher data set is +6.5, reflecting the fact that root-to-node distance generally increased for ArtEx subjects using real states. All of the other data sets exhibited similarly positive median changes in root-to-node distance, with most artificially fossilized extant taxa with real states above the midpoints of their respective areas of possible change given their full-data topologies (Fig. 11).

When their character states were randomized as binary, three of the six data sets (Pattinson, Halliday-All, Halliday-50) exhibited root-to-node change at the midpoints of their expected range, evident by the squares in Figure 11 falling at or near the solid gray line. The other three data sets exhibited a pronounced bias with binary randomizations: root-to-node distance decreased in Huttenlocker and Livezey-Zusi and increased in Asher, with squares below or above, respectively, the solid gray line in Figure 11. For example, rather than moving from its nested position adjacent to Carnivora to *Tamandua* using its real character states, with randomized states *Manis* appeared in many different clades farther from the root than *Tamandua*, ranging from sister to *Elephantulus*, paenungulates, cetaceans, and chiropterans. Some of these clades which attracted the randomized *Manis* exhibited long branches (e.g., chiropterans); others occupied relatively short branches but had the highest proportions of inapplicable characters, such as cetaceans and *Trichechus*. Both of the latter occupied highly nested positions and contributed to increased root-to-node distance when taxa with randomized states were attracted to them.

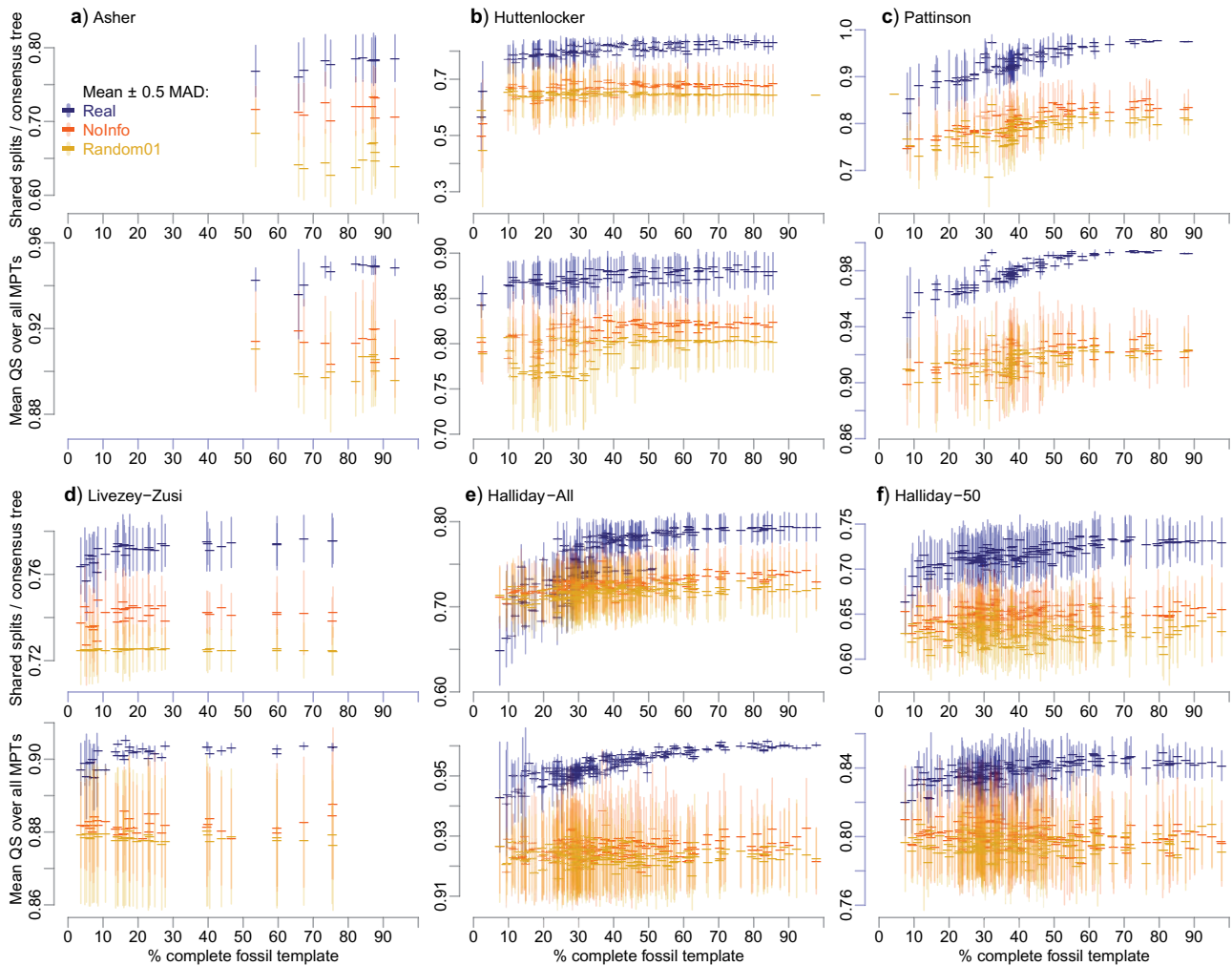


FIGURE 9. Congruence of artificial-extinction topologies with well-corroborated trees. Each horizontal bar shows the mean number of shared splits using strict consensus (top Y-axes) or quartet similarity averaged across all MPTs (bottom Y-axes), averaged across all extant subjects per fossil template. Vertical lines represent one median absolute deviation (0.5 above, 0.5 below) around each mean. Templates vary in completeness for each morphological data set (x-axis). In all cases, characters missing in a fossil template were coded as missing in each extant subject. Remaining characters were coded with the real states in each template (dark shading, “Real”), randomized using states drawn from different extant taxa (middle shading, “NoInfo”), or randomized with states 0 or 1 (light shading, “Random01”).

In nearly all cases in the Livezey-Zusi data set, binary state-randomization (but not real states) resulted in the attraction of the artificial fossil to the root branch (Crocodylia), which was the longest. Decreases in root-to-node distance was therefore pervasive for the Livezey-Zusi data set (Fig. 11a), but only with binary randomized character states due to long-branch attraction.

The four longest branches in Huttenlocker led to *Rattus*, *Erinaceus*, Theria to Marsupialia, Theria to Placentalia, and root to *Ornithorhynchus*. The first two longest branches were within Placentalia and not particularly near the root. However, the latter three were near the root, and the tendency of artificial fossils with binary randomized states to drift toward one of these branches resulted in an overall negative change in root-to-node distance (Fig. 11e).

Size of sister group.—Unlike root-to-node distance, “sisterSize” (TreeTools, Smith 2019a) defines any taxon with a sister group size of one as maximally crownward, regardless of distance from the root node. This has the advantage of representing the directionality of topological change of a given taxon after deletion of non-fossilizable data without being influenced by topological rearrangements that do not immediately pertain to that taxon (Fig. 2). Most taxa in each of our optimal, bifurcating trees derived from our six full data sets (Supplementary Appendix S3 available on Dryad) have only one sister taxon (Fig. 12) and have, accordingly, more potential for “stemward” than “crownward” change when defined using sisterSize.

As evident in the summary panels in Figure 12, taxa made artificially extinct with real states appeared either

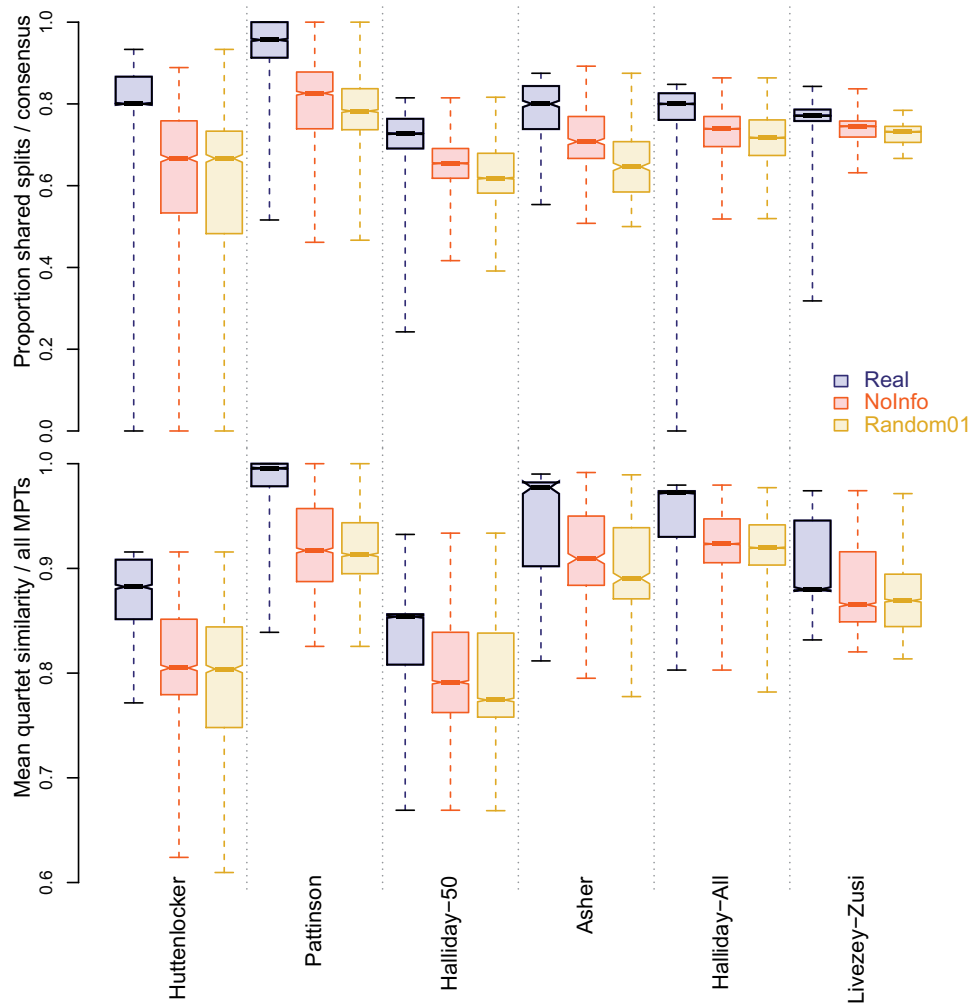


FIGURE 10. Congruence of artificial-extinction topologies with well-corroborated trees for each data set. Each horizontal bar shows the number of shared splits using strict consensus (top Y-axis) or quartet similarity averaged across all MPTs (bottom Y-axis), averaged across all extant subjects per fossil template. Data sets are ordered from largest (left) to smallest (right) difference in median shared splits obtained by real versus 01-randomized character states (see Table 2). Boxes denote median and interquartile range, whiskers the range; non-overlapping notches represent a significant difference in medians (Chambers et al. 1983). In all cases, characters missing in a fossil template were coded as missing in each extant subject. Remaining characters were coded with the real states in each template (dark shading), randomized using states drawn from different extant taxa (middle shading, “NoInfo”), or randomized with states 0 or 1 (light shading, “Random01”).

with slightly fewer mean sister taxa than both “random” and “uncertain” (Asher, Halliday-50), or between these null expectations (Halliday-All, Huttenlocker, Pattinson, Livezey-Zusi). The former corresponds to a slight “crownward” movement, the latter to no significant directional change. With randomized character states, we observed pervasive bias in two cases: Livezey-Zusi (Fig. 12d) and Huttenlocker (Fig. 12b), in which mean sister size greatly increased. This corresponds to “stemward” change, reflected also in the greatly reduced root-to-node distances for subjects with randomized character states (Fig. 11). Unlike the “crownward” change reported for such subjects from the Asher data set based on increased root-to-node distance (Fig. 11f, squares), sister size does not follow suit and instead exhibits a slight increase (Fig. 12a). This underscores the

importance of tree shape in quantifying “crownward” versus “stemward” directional change for a given data set (Fig. 2).

Using real character states, and measured either by root-to-node distance (Fig. 11) or size of sister group (Fig. 12), we find no evidence that the loss of non-fossilizable data results in a systematic bias that causes fossils to appear misleadingly primitive.

DISCUSSION

Our study has so far used interrelationships among extant taxa as a standard against which to assess phylogenetic information content of characters available for fossils. An assumption of this study is that recovery

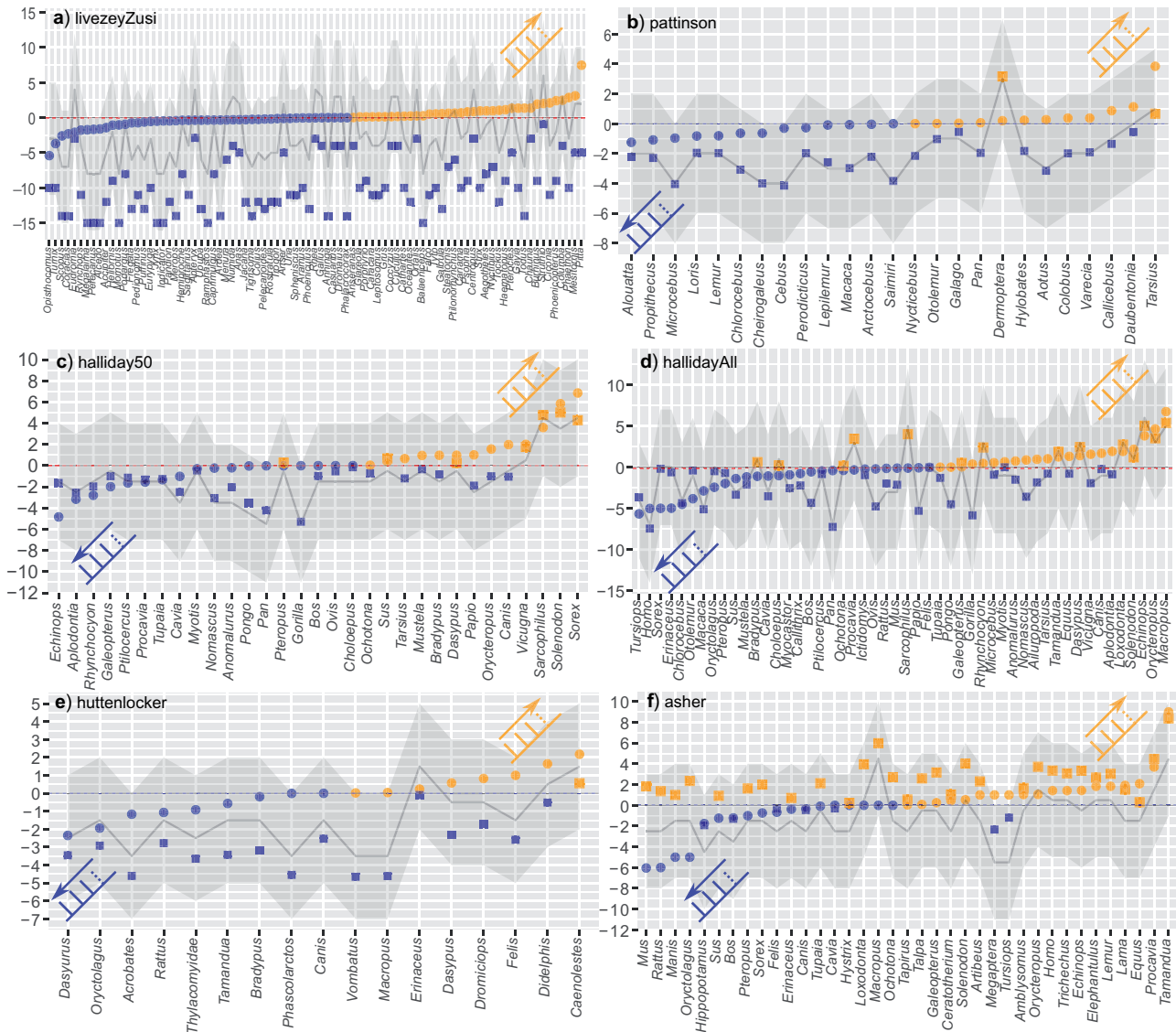


FIGURE 11. Mean change in root-to-node distance (Y-axis) as a result of artificially fossilizing each extant subject taxon (X-axis) using characters from fossil templates, averaging root-to-node distances for all MPTs in each subject-template combination and averaging change for all templates within a subject. As indicated by the arrow insets, light and dark polygons refer to, respectively, mean increase and decrease/no-change in root-to-node distance after artificial extinction. Real character states are shown by circles, binary randomized states with squares. The gray shaded region shows the potential for change in root-to-node distance given the placement of each taxon in each data set's optimal, equally-weighted MP topology of extant taxa, the midpoint of which is shown by the solid gray line.

of well-corroborated splits among extant taxa implies that a given data set can also accurately reconstruct splits among fossil taxa. A strong test of this assumption would come with the discovery of independent phylogenetic data (e.g., molecular) from extinct taxa. Genomic and proteomic data now available for upper Pleistocene fossils have confirmed their morphologically hypothesized associations with, for example, proboscideans (Barnes et al. 2007), perissodactyls (Cappellini et al. 2019), xenarthrans (Delsuc et al. 2019), and hominins (Slon et al. 2018). Many truly ancient, pre-Neogene taxa are much more difficult to place based on their morphology, and molecular data for such taxa are not likely to be forthcoming anytime soon (Saitta et al. 2019).

Our analyses sampling morphology (Fig. 5) and extant taxa known for DNA (Figs. 7 and 8) comprise a weaker, but more accessible, test. If it were true that accurate phylogenetic reconstruction was only possible given availability of molecular data, then increased accuracy should be roughly proportional to the level of DNA sampling. However, our results show that even with no DNA, some congruence exists (Figs. 5 and 8). Furthermore, the addition of a morphological data set to subsampled DNA increases congruence in at least some data sets (Fig. 7), and the addition of taxa known for DNA and morphology can disproportionately increase congruence, for example, any of the points above the red diagonals in Figure 8. In other cases, congruence is below

TABLE 2. Difference in mean proportion shared splits and quartet similarity of topologies generated by real versus binary random character states with the well-corroborated tree for each data set, represented graphically in Fig. 10 and Supplementary Fig. S2 available on Dryad

| Study | Template % complete | Strict MPTs | | Average MPTs | |
|--------------|---------------------|-----------------------------|--------------------|-----------------------------|--------------------|
| | | Real-rand propShared splits | Real-rand quartets | Real-rand propShared splits | Real-rand quartets |
| livezeyZusi | >53 | 0.0550 | 0.0240 | 0.0558 | 0.0241 |
| hallidayAll | >53 | 0.0732 | 0.0399 | 0.0730 | 0.0341 |
| halliday50 | >53 | 0.1042 | 0.0475 | 0.1042 | 0.0469 |
| asher | >53 | 0.1333 | 0.0653 | 0.1225 | 0.0406 |
| pattinson | >53 | 0.1670 | 0.0787 | 0.1722 | 0.0743 |
| huttenlocker | >53 | 0.1824 | 0.0810 | 0.1905 | 0.0811 |
| livezeyZusi | All | 0.0472 | 0.0183 | 0.0511 | 0.0232 |
| hallidayAll | All | 0.0602 | 0.0304 | 0.0659 | 0.0293 |
| halliday50 | All | 0.0955 | 0.0490 | 0.0922 | 0.0448 |
| asher | All | 0.1333 | 0.0653 | 0.1225 | 0.0406 |
| pattinson | All | 0.1478 | 0.0755 | 0.1513 | 0.0670 |
| huttenlocker | All | 0.1608 | 0.0726 | 0.1686 | 0.0700 |

Note: “Strict MPTs” shows values calculated based on a strict consensus across all MPTs for a given subject-template combination; “average MPTs” shows values calculated based on an average across all MPTs for a given subject-template combination. Template % complete shows results from analyses starting with either the minimum complete value from the Asher data set (“>53”) or all templates from all data sets (“all”) which range from 2% to 98% complete (see Supplementary Fig. S1 available on Dryad).

what one would expect given a proportional relationship between number of added taxa with DNA, for example, studies with proportion of shared splits below the red diagonal in Figure 8. As any study approaches complete congruence, the distances of points in Figure 8 above the red diagonal necessarily become smaller. Nonetheless, there are clear differences among data sets (Figs. 5, 7, and 8), indicating that not all morphological data sets retain equally informative phylogenetic signal.

Springer et al. (2007) undertook a variant of the artificial extinction method used here, which they termed “pseudoextinction.” Their analysis emphasized cases in which morphological data could not accurately reconstruct high-level relationships among placental mammals. By deleting molecular data from groups of taxa at the Linnean rank of “Order,” and using the 196-character data set of Asher et al. (2003; similar to the Asher 2007 data set used here) they concluded that “[M]orphological studies of eutherian interordinal relationships have failed to separate homology and homoplasy and have consistently been misled by the latter.” Their results differ from ours in part because they created artificial fossils using the arbitrary rank of Linnean Order, usually entailing deletion of DNA from multiple extant taxa per analysis (rather than just one genus per analysis as in our Fig. 9). This is comparable to a subset of our analyses in which multiple taxa known for DNA were deleted simultaneously (e.g., lower levels of taxon sampling in our Fig. 8).

Our evaluation of individual morphological data sets with little or no added DNA also shows incongruence with well-corroborated parts of the phylogenetic tree (Figs. 5, 7, and 8). However, the extent of this incongruence varies, and underscores our interpretation that “morphology” should not be judged as a whole as either “good” or “bad.” Rather, different morphological data sets retain at least some phylogenetic signal depending on the available sample of taxa and characters, and

data sets not sampled here will exhibit their own idiosyncrasies. In practice, phylogenetic questions often include groups (e.g., fossil adapiforms, microbiotherians, paenungulates, ratites) that are nested among extant groups (e.g., primates, marsupials, afrotherians, paleognaths). Such groups are amenable to analyses of combined morphology plus densely sampled DNA alignments. Since extant taxa can be accurately reconstructed using only morphological characters when they are sampled along with others known for both DNA and hard-tissue morphology (Fig. 9), we hypothesize that this is also true for fossils. Clearly, this becomes more difficult with either fewer taxa known for DNA (e.g., for long-extinct clades without close extant relatives) or when a data set is missing the morphological characters, present in fossils, that would otherwise enable phylogenetic reconstruction in a combined-data context (e.g., Halliday-All with many extant taxa incompletely coded for morphology). Regarding the latter point, our results support the interpretation of Guillerme and Cooper (2016a,b) that coding morphological data for living taxa improves phylogenetic accuracy.

Discriminating Among Morphological Data Sets

The data set for primates (Seiffert et al. 2009; Pattinson et al. 2015) was the only one to recover all of its well-corroborated nodes with increasing DNA subsamples (Figs. 7 and 8) and fossil template completeness (Fig. 9). The addition of morphological data increases congruence in all of the subsampled DNA alignments (Fig. 7), as does the use of fossil templates of increasing morphological completeness (Figs. 9 and 10). We would argue that the good performance of the Pattinson data set is due to the fact that primates represent a younger radiation that has left behind numerous fossils with characters that are relatively more informative than

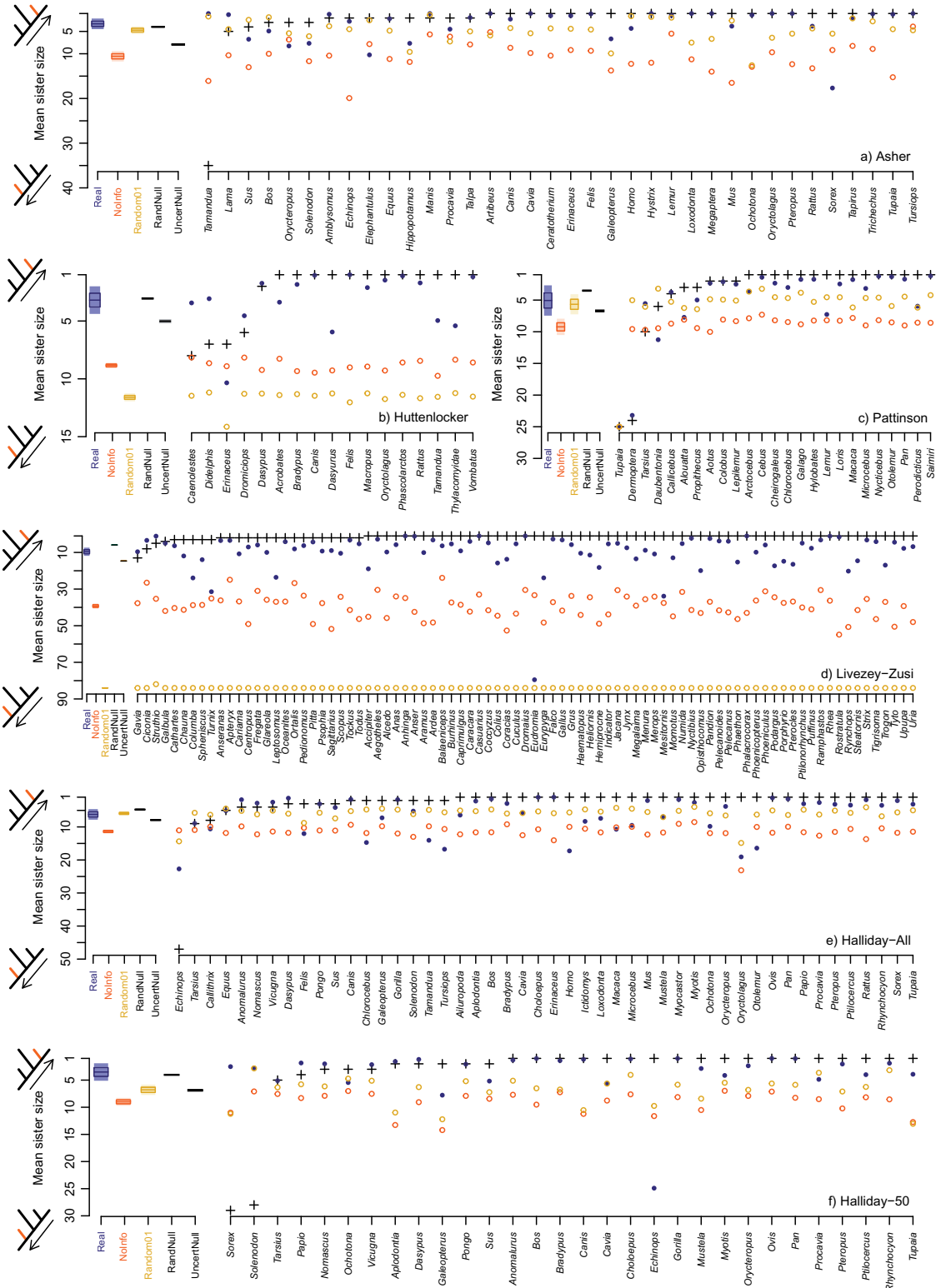


FIGURE 12. Mean sister group size of artificially fossilized subjects in all most parsimonious topologies, averaged across all fossil templates. Plus (“+”) denotes sister-group size for given taxon in topology reconstructed from the complete, original data set; sister group size given by solid, dark-shaded dots represent artificially fossilized taxa with real character states (“Real”), open, medium-shaded circles randomized states drawn from other extant taxa (“Nolnfo”), open, light-shaded circles binary randomized states (“Random01”). Panel at left summarizes sister-group size based on real and randomized states for all taxa in each data set, along with two null models: “random null” (moving the subject to a random edge on the tree) and “uncertainty null” (moving the subject to an edge with a probability proportional to the degree of uncertainty in character state at that edge).

TABLE 3. Optimal topologies from this study compared with those figured by Asher (2007, Fig. 3), Halliday et al. (2019, Figs. 1 and 2), and Huttenlocker et al. (2018: Supplementary Fig. 9A available on Dryad), as indicated in “source” column, and given in newick format in Supplementary Appendix S3 available on Dryad

| Data Set | Data | Taxa | Length | MPTs | Optimality | Source | Tree name | RF | QS | Shared | WC splits |
|---------------|-------|------|--------------|------|------------|----------|----------------|----|------|--------|-----------|
| asher | comb | e+f | 94,175 | 34 | MP | majrules | asherMR | 11 | 0.98 | 27 | 30 |
| asher | comb | e+f | 94,175 | 34 | MP | strict | asherStrict | 13 | 0.74 | 22 | 30 |
| asher | comb | e | 94,029 | 3 | MP | strict | asherExtant | 9 | 0.98 | 27 | 30 |
| asher | comb | e+f | — | — | Bayes | fig3 | asher2007F3 | 9 | 0.99 | 28 | 30 |
| asher* | comb | e+f | 4258.25142 | 1 | MP k = 8 | best | asherK8 | 9 | 0.98 | 28 | 30 |
| halliday | comb | e+f | 121,989 | 8 | MP | majrules | hallMR | 16 | 0.97 | 38 | 44 |
| halliday | comb | e+f | 121,989 | 8 | MP | strict | hallStrict | 16 | 0.97 | 38 | 44 |
| halliday | comb | e | 114,942 | 1 | MP | best | hallExtant | 16 | 0.97 | 38 | 44 |
| halliday | comb | e+f | — | — | likelihood | fig1 | hall2019F1 | 14 | 0.86 | 36 | 43 |
| halliday | comb | e | — | — | likelihood | fig2 | hall2019F2 | 4 | 1 | 34 | 34 |
| halliday* | comb | e+f | 11,026.82414 | 1 | MP k = 2 | best | hallidayAllK2 | 10 | 0.98 | 41 | 44 |
| halliday50 | comb | e+f | 85,573 | 276 | MP | majrules | hall50MR | 13 | 0.86 | 21 | 26 |
| halliday50 | comb | e+f | 85,573 | 276 | MP | strict | hall50Strict | 13 | 0.86 | 21 | 26 |
| halliday50 | comb | e | 78,548 | 1 | MP | best | hall50Extant | 13 | 0.86 | 21 | 26 |
| halliday50* | comb | e+f | 46.99872 | 1 | MP k = 999 | best | halliday50K999 | 13 | 0.95 | 21 | 26 |
| huttenlocker | comb | e+f | 50,854 | 640 | MP | majrules | huttMR | 4 | 0.92 | 13 | 15 |
| huttenlocker | morph | e+f | 2667 | 1000 | MP-morph | strict | huttmorphS | 12 | 0.76 | 9 | 15 |
| huttenlocker* | comb | e+f | 50,854 | 640 | MP | strict | huttStrict | 4 | 0.92 | 13 | 15 |
| huttenlocker | comb | e | 48,993 | 1 | MP | best | huttExtant | 4 | 0.92 | 13 | 15 |
| huttenlocker | morph | e+f | — | — | Bayes | suppF9a | hutt2018sF9a | 13 | 0.76 | 8 | 15 |
| livezeyZusi | comb | e+f | 995,720 | 2 | MP | majrules | livezeyMR | 33 | 0.88 | 60 | 70 |
| livezeyZusi | comb | e+f | 995,720 | 2 | MP | strict | livezeyStrict | 33 | 0.88 | 60 | 70 |
| livezeyZusi | comb | e | 992,701 | 1 | MP | best | livezeyExtant | 33 | 0.88 | 60 | 70 |
| livezeyZusi* | comb | e+f | 65,443.97516 | 1 | MP k = 4 | best | livezeyZusiK4 | 25 | 0.99 | 64 | 70 |
| pattinson | comb | e+f | 47,875 | 460 | MP | majrules | pattMR | 0 | 1 | 23 | 23 |
| pattinson* | comb | e+f | 47,875 | 460 | MP | strict | pattStrict | 0 | 1 | 23 | 23 |
| pattinson | comb | e | 46,010 | 1 | MP | best | pattExtant | 0 | 1 | 23 | 23 |

Note: Those shown with an asterisk are depicted in Supplementary Figures S3–S8 available on Dryad. “Comb” and “morph” indicate combined DNA + morphology or just morphology for a given analysis, “e” and “f” = extant only or fossil + extant taxa, length and MPTs = length and number of most parsimonious trees, “k” = implied weighting concavity value, QS = quartet similarity, RF = Robinson–Foulds distance, treename = ID given in Supplementary Appendix S3 available on Dryad, shared = number of splits shared with well-corroborated tree, WC splits = maximum possible splits shared with well-corroborated tree (modified for Halliday following taxon sample in Halliday et al. (2019)). MP = maximum parsimony; “strict”, “majrules”, and “best” indicate, respectively, strict, majority-rules, and single-best MP topology.

those from our bird, Mesozoic and early Paleogene mammal data sets. All of our other data sets focus on at least some clades that are substantially older than those sampled by Pattinson et al. (2015).

In terms of the impact of morphological data, the most problematic data set in our study is that of Halliday et al. (2019). The abundance of incompletely coded, extant taxa in Halliday-All made accurate reconstruction of well-corroborated splits highly variable, extensively overlapping with ArtEx topologies generated by randomized states (Figs. 4, 9e, and 10) and more dependent on inclusion of DNA from living taxa to achieve congruence compared to other studies. Our deletion of the 19 extant subject taxa missing 50% or more morphological characters (Halliday-50) increased the performance of their real character states over randomized ones (Figs. 4 and 10; Table 2). However, it made overall congruence worse due to the inability of parsimony to accurately root the Halliday-50 sample of placentals near the Afrotherian-Xenarthran clade (Tarver et al. 2016).

Calculating congruence by averaging across all MPTs for a given subject-template combination, rather than using a strict consensus thereof, decreased the variation in both Halliday-All and Halliday-50. This shows that a set of optimal trees can still retain some areas of agreement with an independent topology, masked when

they are summarized with a blunt instrument such as the strict consensus. However, strict consensus topologies remain widely used in paleontological systematics, and intuitively convey when results from a given data set are ambiguous (or not). We would predict that once the many missing entries for extant taxa in their data set are coded, Halliday-All will improve in its capacity to consistently recover known parts of the mammalian tree.

The Livezey-Zusi data set shows mixed results. On one hand, it shows only a small difference in congruence achieved by real versus randomized character states (Fig. 10), and it exhibits little improvement in congruence beyond templates that are around 20% complete (Figs. 9c,f). The addition of morphology reduced the number of splits shared with the well-corroborated tree at some levels of DNA sampling but not others, particularly as measured by quartet similarity (Fig. 7). Mayr (2008) criticized many of the morphological codings of Livezey-Zusi, underscoring the observation that Aves is a challenging radiation to resolve. It has ancient roots in the Mesozoic and comes from a morphologically and ecologically distinct stem lineage (Dinosauria). Furthermore, Livezey-Zusi has the greatest proportion of fragmentary templates, with most (22 of 39) missing 80% or more of their morphological characters (Supplementary Fig. S1 available on Dryad).

None of the 12 templates for Asher are in this range; there are 18 of 99 in Huttenlocker, 6 of 85 in Pattinson, and 18 of 174 in Halliday-All and Halliday-50. The high frequency of missing data among Livezey-Zusi fossil templates, and its anatomically and ecologically distant root taxon Crocodylia, likely further reduces the capacity of the Livezey-Zusi data set to reconstruct well-corroborated nodes. Compared to Halliday-All, there is therefore less reason to assert that further coding of extant Aves for morphology will increase the ability of this data set to contribute to phylogenetic accuracy. Instead, increasing the completeness of fossil templates, along with further resolution of the Neoavian comb (which remains the least well-resolved of our well-corroborated trees; Fig. 1) seem promising avenues for further work.

Parsimony Versus Published Bayesian and Likelihood Trees

Our use of equal-weights parsimony is a practical one. As noted in Materials and Methods, running many thousands of individual phylogenetic analyses required that each one find an optimal tree within seconds or minutes. While it is fast, well-known drawbacks of parsimony include susceptibility to long-branch artifacts, likely responsible for the suboptimal asymptotes of some of the subsampled DNA alignments (Fig. 7) and ArtEx experiments (Fig. 9). While a comprehensive analysis of competing optimality criteria is beyond the scope of our study, we can still investigate the performance of our optimal parsimony topologies with Bayesian and likelihood topologies already published for four of our six data sets.

Pattinson and Asher.—The optimal parsimony topology for the Pattinson data set under equal and all implied weights (Supplementary Fig. S3 available on Dryad) shares all well-corroborated nodes with the likelihood analysis of extant taxa figured by Springer et al. (2012: Fig. 1), and with the topology generated by the same morphology-DNA alignment analyzed with a Bayesian optimality criterion depicted by Pattinson et al. (2015: Fig. 1). The Bayesian topology from Asher (2007) was based on a smaller DNA alignment (17kb including indels) than that used here, and it sampled only three (*Ukhaatherium*, *Zalambdalestes*, *Centetodon*) of 12 fossil templates. It shows slightly more congruence with the well-corroborated tree (28 of 30 or 0.93 shared splits) by placing *Solenodon* as sister to *Talpa-Erinaceus-Sorex*, rather than *Talpa-Solenodon* as shown in the equally-weighted parsimony topology derived from the combined Asher-Upham data set. Optimal implied weights applied to the Asher-Upham data set resembled the Asher (2007: Fig. 3) Bayesian tree regarding *Solenodon*, but differed in resolving *Tupaia*-dermopterans closer to Glires than primates (Supplementary Fig. S4 available on Dryad).

Halliday.—The likelihood topology including extant taxa and fossils from Halliday et al. (2019, Fig. 1) shows less congruence (0.84 proportion shared splits) than

our equally-weighted (0.86) and best implied-weights (0.93) parsimony topologies for their morphology data set combined with the Upham DNA alignment (see Table 3 and Supplementary Fig. S5 available on Dryad). This is due to the placement by Halliday et al. (2019, Fig. 1) of the rodents *Aplodontia* and *Anomalurus* closer to cavioids than to (respectively) sciuroids and muroids, of Dermoptera closer to tupaiids than primates, and of the macroscelidid *Rhynchocyon* closer to *Procavia* than *Loxodonta*. Figure 2 from Halliday et al. (2019) excludes all fossils as well as the extant taxa with incongruent placements in their Figure 1 (*Aplodontia*, *Anomalurus*, the dermopteran and *Rhynchocyon*). For the taxa they had in common with our well-corroborated tree (Fig. 1), Figure 2 of Halliday et al. (2019) was completely congruent, differing only in resolving some of the polytomies shown in our well-corroborated tree (such as *Tupaia*-primates, *Cavia* at the rodent root, and perissodactyls-carnivorans; Halliday et al. 2019: Fig. 2 did not include *Manis* or *Orycteropus*).

Huttenlocker.—The Bayesian topology of morphological data from Huttenlocker et al. (2018: their Supplementary Fig. 9a available on Dryad) shows less congruence (0.53 proportion shared splits) than equal- or implied-weights parsimony applied to their data set plus the Upham et al. (2019) alignment (0.87 proportion shared splits for a majority rules consensus; see Table 3 and Supplementary Fig. S6 available on Dryad). Huttenlocker et al. (2018) depict a number of incongruent nodes in their Bayesian tree of morphological data, such as *Erinaceus* and murids as the two basal-most placental branches, *Macropus* closer to Vombatiformes than *Acrobates*, the peramelemorphian taxon Thylacomyidae closer to diprotodontians than *Dasyurus*, and *Didelphis* at the base of Marsupialia rather than *Caenolestes*. Due to a polytomy in Xenarthra, their Bayesian topology also shows less congruence than MP applied to their morphological data alone including both extant taxa and fossils. Parsimony resolves [(*Bradypus*, *Tamandua*), *Dasyurus*], as evident in both their (Huttenlocker et al. 2018: their Supplementary Fig. 9b available on Dryad) and our parsimony analyses of their morphological data (Table 3; Supplementary Fig. S6 available on Dryad).

Despite its well-documented susceptibility to long-branch effects (e.g., Swofford et al. 2001), our results show that parsimony applied to combined morphology-DNA data sets results in greater congruence than a Bayesian optimality criterion applied to morphology alone (Huttenlocker et al. 2018). Applied to the morphological data set of Halliday et al. (2019) combined with DNA from Upham et al. (2019), parsimony resulted in greater congruence than the majority-rule likelihood topology of the morphology-DNA data of Halliday et al. (2019, Fig. 1). In addition to its computational speed, parsimony as an optimality criterion benefits from not assuming the same rates of change for all characters proportionate to a given branch length (Goloboff et al. 2018). Furthermore, when taking into account resolution

and using implied weights, its performance in at least some cases is comparable to that of Bayesian consensus of post-burnin topologies (Smith 2019c), and compared to Bayesian methods it shows more correspondence with the tetrapod stratigraphic record (Sansom et al. 2018).

Directional Bias Caused by Fossilization

Our results contradict the claim that fossils systematically appear closer to the root taxon than they actually are. Instead, taking into account the shape of each full data set topology, most directional change observed after artificial fossilization was toward higher root-to-node distances (Fig. 11) and similar or smaller sister group sizes compared to null models (Fig. 12), both of which contradict the hypothesis of pervasive “stemward” directional change. Only when character states in artificial fossils were randomized did we identify some examples of pervasive bias (e.g., Fig. 10a). Loss of data in fossils is ubiquitous, but as long as their character states are accurately scored, “stemward slippage” as a result of this loss is not.

CONCLUSIONS

Results in this study are pertinent to a number of issues regarding the phylogenetic analysis of fossils. Among the most important is the substantial increase in congruence with known parts of the Tree of Life when fossilizable, morphological data are analyzed simultaneously with DNA from living taxa. This alone provides the single greatest improvement in congruence we have observed among our data sets. The most congruent topologies from analyses using morphology alone (Fig. 5) are still worse than the least congruent topologies joining morphology and DNA (Fig. 7). Although the addition of morphological characters does not always increase congruence when combined with DNA (Fig. 7), overall accuracy is still better in a combined data context than with morphology alone (Table 3). Therefore, our take-home message for anyone interested in reconstructing the positions of fossils on the Tree of Life is, whenever possible, include DNA and morphology from extant taxa simultaneously with morphology from fossils.

It is counterintuitive to expect the addition of DNA from extant taxa to alter relationships among fossils, or for small morphological partitions to influence placement of extant taxa sampled for much larger molecular partitions, but such effects can happen (Asher et al. 2005; Wiens et al. 2010). Our optimal parsimony topologies from each data set (Supplementary Figs. S3–S8 available on Dryad) often support the resolved parts of the studies in which each morphological data set was first published, with some exceptions. For example, our analysis of Huttenlocker et al. (2018), including DNA for living taxa shared with Upham et al. (2019), supports their interpretation that Triassic and Jurassic haramiyids are not crown mammals. However, the

Paleocene metatherians *Andinodelphys* and *Pucadelphys* do not appear as successively distant sister taxa to Marsupialia (Huttenlocker et al. 2018: Supplementary Fig. 9 available on Dryad), but in our equal- and most implied-weights analyses are within crown Marsupialia as the sister taxon of didelphids (Supplementary Fig. S6 available on Dryad). Our analysis of Halliday-All (Supplementary Fig. S5 available on Dryad) and Halliday-50 (Supplementary Fig. S8 available on Dryad) resolved one or more groups of fossil plesiadapiforms closer to tupaiids or glires, rather than primates as figured by Halliday et al. (2019, Fig. 1). The analysis of Livezey-Zusi (2007: Fig. 12) reconstructs *Hesperornis* and *Ichthyornis* closer to *Lithornis*-Neornithes than *Apsaravis*, in contrast to our optimal implied-weights analysis showing *Apsaravis* closer to *Lithornis*-Neornithes (Supplementary Fig. S7 available on Dryad). We do not claim that any of these relationships are necessarily accurate, which would require a dedicated analysis (including additional phylogenetic reconstruction techniques) in each case. However, several illustrate the capacity of molecular data from living taxa to influence the phylogenetic placement of fossils.

Vertebrate biologists now have the luxury of understanding many well-corroborated branches for living taxa on the phylogenetic Tree of Life. Over the past two decades, this tree has revolutionized how biologists understand evolution. Fossils will never have as much data as their living descendants, but thanks to our understanding of living species, it may yet be possible to obtain more confidence in the phylogenetic affinities of long extinct species. Paleontologists should take advantage of this current knowledge of extant species' phylogeny when attempting to reconstruct the evolutionary history of fossils.

SUPPLEMENTARY MATERIAL

Data available from the Dryad Digital Repository: <https://dx.doi.org/10.5061/dryad.w3r2280q3>.

ACKNOWLEDGMENTS

We are grateful to Robin Beck, Seraina Klopstein, Aime Rankin, Robert Sansom, and Oscar Wilson for discussion, Sophia Rodrigues for assistance assembling the morphological matrix of Livezey and Zusi (2006), and Daniel Field for advice on avian phylogenetics. R.J.A. thanks Seraina Klopstein for writing the “node height” function to help calculate root-to-node distances (Supplementary Table S5 available on Dryad), and his department (Zoology) and college (Trinity Hall) at the University of Cambridge for sabbatical time that facilitated this research. Both authors acknowledge sponsorship of the Willi Hennig Society that enabled our use of TNT (Goloboff and Catalano (2016)) and are grateful to three anonymous reviewers and editors Liliana Dávalos and Bryan Carstens for their comments

which have improved our manuscript. The authors declare no financial or commercial conflicts of interest with the content of this article.

REFERENCES

- Asher R.J. 2007. A web-database of mammalian morphology and a reanalysis of placental phylogeny. *BMC Evol. Biol.* 7:108.
- Asher R.J., Bennett N., Lehmann T. 2009. The new framework for understanding placental mammal evolution. *BioEssays*. 31:853–864.
- Asher R.J., Emry R.J., Mckenna M.C. 2005. New material of *Centetodon* (Mammalia, Lipotyphla) and the importance of (missing) DNA sequences in systematic paleontology. *J. Vertebr. Palaeontol.* 25:911–923.
- Asher R.J., Geisler J.H., Sánchez-Villagra M.R. 2008. Morphology, paleontology, and placental mammal phylogeny. *Syst. Biol.* 57:311–317.
- Asher R.J., Hofreiter M. 2006. Tenrec phylogeny and the noninvasive extraction of nuclear DNA. *Syst. Biol.* 55:181–194.
- Asher R.J., Novacek M.J., Geisler J.H. 2003. Relationships of endemic African mammals and their fossil relatives based on morphological and molecular evidence. *J. Mamm. Evol.* 10:131–164.
- Asher R.J., Smith M.R., Rankin A., Emry R. 2019. Congruence, fossils, and the evolutionary tree of rodents and lagomorphs. *R. Soc. Open Sci.* 6:190387.
- Barnes I., Shapiro B., Lister A., Kuznetsova T., Sher A., Guthrie D., Thomas M.G. 2007. Genetic structure and extinction of the woolly mammoth, *Mammuthus primigenius*. *Curr. Biol.* 17:1072–1075.
- Baron M.G., Norman D.B., Barrett P.M. 2017. A new hypothesis of dinosaur relationships and early dinosaur evolution. *Nature*. 543:501.
- Beck R.M.D., Baillie C. 2018. Improvements in the fossil record may largely resolve current conflicts between morphological and molecular estimates of mammal phylogeny. *Proc. R. Soc. B Biol. Sci.* 285:20181632.
- Beck R.M.D., Lee M.S.Y. 2014. Ancient dates or accelerated rates? Morphological clocks and the antiquity of placental mammals. *Proc. R. Soc. Lond. B Biol. Sci.* 281:20141278.
- Benton M.J., Donoghue P.C.J., Asher R.J. 2009. Calibrating and constraining molecular clocks. In: Kumar S., Hedges S.B., editors. *The Timetree of Life*. Oxford: Oxford University Press. p. 35–86.
- Benton M.J., Donoghue P.C.J., Asher R.J., Friedman M., Near T.J., Vinther J. 2015. Constraints on the timescale of animal evolutionary history. *Palaeontologica Electronica* 18.1.1FC:1–106.
- Briggs D.E.G. 1995. Experimental taphonomy. *Palaios*. 10:539–550.
- Brochu C.A. 2003. Phylogenetic approaches toward crocodylian history. *Annu. Rev. Earth Planet. Sci.* 31:357–397.
- Brochu C.A. 1997. Morphology, fossils, divergence timing, and the phylogenetic relationships of *Gavialis*. *Syst. Biol.* 46:479–522.
- Cappellini E., Welker F., Pandolfi L., Ramos-Madrigal J., Samodova D., Rüter P.L., Fotakis A.K., Lyon D., Moreno-Mayar J.V., Bukhshianidze M., Jersie-Christensen R.R. 2019. Early Pleistocene enamel proteome from Dmanisi resolves *Stephanorhinus* phylogeny. *Nature*. 574:103–107.
- Chambers J.M., Cleveland W.S., Kleiner B., Tukey P.A. 1983. Graphical methods for data analysis. Boston (MA): Wadsworth & Brooks/Cole.
- Delsuc F., Kuch M., Gibb G.C., Karpinski E., Hackenberger D., Szpak P., Martínez J.G., Mead J.I., McDonald H.G., MacPhee R.D.E., Billet G. 2019. Ancient mitogenomes reveal the evolutionary history and biogeography of sloths. *Curr. Biol.* 29:2031–2042.
- Donoghue P.C.J., Purnell M.A. 2009. Distinguishing heat from light in debate over controversial fossils. *BioEssays*. 31:178–189.
- dos Reis M., Inoue J., Hasegawa M., Asher R.J., Donoghue P.C.J., Yang Z. 2012. Phylogenomic datasets provide both precision and accuracy in estimating the timescale of placental mammal phylogeny. *Proc. R. Soc. Lond. B Biol. Sci.* 279:3491–3500.
- Eernisse D.J., Kluge A.G. 1993. Taxonomic congruence versus total evidence, and amniote phylogeny inferred from fossils, molecules, and morphology. *Mol. Biol. Evol.* 10:1170–1195.
- 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.
- Felsenstein J. 1978. Cases in which parsimony or compatibility methods will be positively misleading. *Syst. Zool.* 27:401–410.
- Flannery Sutherland J.T., Moon B.C., Stubbs T.L., Benton M.J. 2019. Does exceptional preservation distort our view of disparity in the fossil record? *Proc. R. Soc. B*. 286:20190091.
- Foley N.M., Springer M.S., Teeling E.C. 2016. Mammal madness: is the mammal tree of life not yet resolved? *Philos. Trans. R. Soc. B*. 371:20150140.
- Gatesy J., Meredith R.W., Janecka J.E., Simmons M.P., Murphy W.J., Springer M.S. 2017. Resolution of a concatenation/coalescence kerfuffle: partitioned coalescence support and a robust family-level tree for Mammalia. *Cladistics*. 33:295–332.
- 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.
- Gatesy J., O'Leary M.A. 2001. Deciphering whale origins with molecules and fossils. *Trends Ecol. Evol.* 16:562–570.
- Goloboff P.A., Carpenter J.M., Arias J.S., Esquivel D.R.M. 2008. Weighting against homoplasy improves phylogenetic analysis of morphological data sets. *Cladistics*. 24:758–773.
- Goloboff P.A., Catalano S.A. 2016. TNT version 1.5, including a full implementation of phylogenetic morphometrics. *Cladistics*. 32:221–238.
- Goloboff P.A., Torres Galvis A., Arias J.S. 2018. Parsimony and model-based phylogenetic methods for morphological data: comments on O'Reilly et al. *Palaeontology*. 61:625–630.
- Guillerme T., Cooper N. 2016a. Effects of missing data on topological inference using a total evidence approach. *Mol. Phylogenet. Evol.* 94:146–158.
- Guillerme T., Cooper N. 2016b. Assessment of available anatomical characters for linking living mammals to fossil taxa in phylogenetic analyses. *Biol. Lett.* 12:20151003.
- Halliday T.J.D., dos Reis M., Tamuri A.U., Ferguson-Gow H., Yang Z., Goswami A. 2019. Rapid morphological evolution in placental mammals post-dates the origin of the crown group. *Proc. R. Soc. B*. 286:20182418.
- Huttenlocker A.K., Grossnickle D.M., Kirkland J.I., Schultz J.A., Luo Z.-X. 2018. Late-surviving stem mammal links the lowermost Cretaceous of North America and Gondwana. *Nature*. 558:108.
- Jenner R.A. 2004. Accepting partnership by submission? Morphological phylogenetics in a molecular millennium. *Syst. Biol.* 53:333–342.
- Klompaker A.A., Portell R.W., Frick M.G. 2017. Comparative experimental taphonomy of eight marine arthropods indicates distinct differences in preservation potential. *Palaeontology*. 60:773–794.
- Klopfstein S., Vilhelmsen L., Ronquist F. 2015. A nonstationary Markov model detects directional evolution in hymenopteran morphology. *Syst. Biol.* 64:1089–1103.
- Koch N.M., Parry L.A. 2020. Death is on our side: paleontological data drastically modify phylogenetic hypotheses. *Syst. Biol.* 69:1052–1067.
- Krause D.W., Hoffmann S., Hu Y., Wible J.R., Rougier G.W., Kirk E.C., Groenke J.R., Rogers R.R., Rossie J.B., Schultz J.A., Evans A.R. 2020. Skeleton of a Cretaceous mammal from Madagascar reflects long-term insularity. *Nature*. 581:421–427.
- Lee M.S.Y. 2001. Molecules, morphology, and the monophyly of diapsid reptiles. *Contrib. to Zool.* 70:1–22.
- Lee M.S.Y., Camens A.B. 2009. Strong morphological support for the molecular evolutionary tree of placental mammals. *J. Evol. Biol.* 22:2243–2257.
- Lee M.S.Y., Palci A. 2015. Morphological phylogenetics in the genomic age. *Curr. Biol.* 25:R922–R929.
- Livezey B.C., Zusi R.L. 2006. Phylogeny of Neornithes. *Bull. Carnegie Museum Nat. Hist.* 37:1–544.
- Livezey B.C., Zusi R.L. 2007. Higher-order phylogeny of modern birds (Theropoda, Aves: Neornithes) based on comparative anatomy. II. Analysis and discussion. *Zool. J. Linn. Soc.* 149:1–95.

- Mayr G. 2008. Avian higher-level phylogeny: well-supported clades and what we can learn from a phylogenetic analysis of 2954 morphological characters. *J. Zool. Syst. Evol. Res.* 46:63–72.
- Murdock D.J.E., Gabbott S.E., Mayer G., Purnell M.A. 2014. Decay of velvet worms (Onychophora), and bias in the fossil record of lobopodians. *BMC Evol. Biol.* 14:1–10.
- Nanglu K., Caron J.-B., Cameron C.B. 2015. Using experimental decay of modern forms to reconstruct the early evolution and morphology of fossil enteropneusts. *Paleobiology.* 41:460–478.
- O’Leary M.A., Bloch J.I., Flynn J.J., Gaudin T.J., Giallombardo A., Giannini N.P., Goldberg S.L., Kraatz B.P., Luo Z.-X., Meng J., Ni X. 2013. The placental mammal ancestor and the post-K-Pg radiation of placentals. *Science.* 339:662–667.
- Paradis E. 2012. *Analysis of phylogenetics and evolution with R.* 2nd ed. New York: Springer.
- Parks S.L., Goldman N. 2014. Maximum likelihood inference of small trees in the presence of long branches. *Syst. Biol.* 63:798–811.
- Pattinson D.J., Thompson R.S., Piotrowski A.K., Asher R.J. 2015. Phylogeny, paleontology, and primates: do incomplete fossils bias the tree of life? *Syst. Biol.* 64:169–186.
- Prum R.O., Berv J.S., Dornburg A., Field D.J., Townsend J.P., Lemmon E.M., Lemmon A.R. 2015. A comprehensive phylogeny of birds (Aves) using targeted next-generation DNA sequencing. *Nature.* 526:569.
- Pyron R.A. 2015. Post-molecular systematics and the future of phylogenetics. *Trends Ecol. Evol.* 30:384–389.
- Reddy S., Kimball R.T., Pandey A., Hosner P.A., Braun M.J., Hackett S.J., Han K.-L., Harshman J., Huddleston C.J., Kingston S., Marks B.D. 2017. Why do phylogenomic data sets yield conflicting trees? Data type influences the avian tree of life more than taxon sampling. *Syst. Biol.* 66:857–879.
- Reisz R.R., Müller J. 2004. Molecular timescales and the fossil record: a paleontological perspective. *Trends Genet.* 20:237–241.
- Ronquist F., Klopfstein S., Vilhelmsen L., Schulmeister S., Murray D.L., Rasnitsyn A.P. 2012a. A total-evidence approach to dating with fossils, applied to the early radiation of the Hymenoptera. *Syst. Biol.* 61:973–999.
- Ronquist F., Teslenko M., Van Der Mark P., Ayres D.L., Darling A., Höhna S., Larget B., Liu L., Suchard M.A., Huelsenbeck J.P. 2012b. MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. *Syst. Biol.* 61:539–542.
- Saitta E.T., Liang R., Lau M.C.Y., Brown C.M., Longrich N.R., Kaye T.G., Novak B.J., Salzberg S.L., Norell M.A., Abbott G.D., Dickinson M.R. 2019. Cretaceous dinosaur bone contains recent organic material and provides an environment conducive to microbial communities. *Elife.* 8:e46205.
- Sand A., Holt M.K., Johansen J., Brodal G.S., Mailund T., Pedersen C.N. 2014. tqDist: a library for computing the quartet and triplet distances between binary or general trees. *Bioinformatics.* 30(14):2079–2080.
- Sansom R.S. 2016. Preservation and phylogeny of Cambrian ecdysozoans tested by experimental decay of *Priapulid*. *Sci. Rep.* 6:1–12.
- Sansom R.S., Wills M.A. 2013. Fossilization causes organisms to appear erroneously primitive by distorting evolutionary trees. *Sci. Rep.* 3:2545.
- Sansom R.S., Wills M.A., Williams T. 2017. Dental data perform relatively poorly in reconstructing mammal phylogenies: morphological partitions evaluated with molecular benchmarks. *Syst. Biol.* 66:813–822.
- Schliep K.P. 2011. phangorn: phylogenetic analysis in R. *Bioinformatics.* 27:592–593.
- Seiffert E.R., Perry J.M.G., Simons E.L., Boyer D.M. 2009. Convergent evolution of anthropoid-like adaptations in Eocene adapiform primates. *Nature.* 461:1118–1121.
- Slon V., Mafessoni F., Vernot B., de Filippo C., Grote S., Viola B., Hajdinjak M., Peyrégne S., Nagel S., Brown S., Douka K. 2018. The genome of the offspring of a Neanderthal mother and a Denisovan father. *Nature.* 561:113–116.
- Smith A.B. 1998. What does palaeontology contribute to systematics in a molecular world? *Mol. Phylogenet. Evol.* 9:437–447.
- Smith, M.R. 2019a. TreeTools: create, modify and analyse phylogenetic trees. *Comprehensive R Archive Network.* doi: 10.5281/zenodo.3522725.
- Smith, M.R. 2019b. Quartet: comparison of phylogenetic trees using quartet and split measures. *Comprehensive R Archive Network.* doi:10.5281/zenodo.2536318.
- Smith, M.R. 2019c. Bayesian and parsimony approaches reconstruct informative trees from simulated morphological datasets. *Biol. Lett.* 15:20180632. doi:10.1098/rsbl.2018.0632.
- Smith, M.R. 2020. TreeDist: Distances between Phylogenetic Trees. R package version 2.0.3. *Comprehensive R Archive Network.* doi:10.5281/zenodo.3528124.
- Springer M.S., Burk-Herrick A., Meredith R., Eizirik E., Teeling E., O’Brien S.J., Murphy W.J. 2007. The adequacy of morphology for reconstructing the early history of placental mammals. *Syst. Biol.* 56:673–684.
- Springer M.S., Foley N.M., Brady P., Gatesy J., Murphy W. 2019. Evolutionary models for the diversification of placental mammals across the KPg boundary. *Front. Genet.* 10:1241.
- Springer M.S., Meredith R.W., Gatesy J., Emerling C.A., Park J., Rabosky D.L., Stadler T., Steiner C., Ryder O.A., Janečka J.E., Fisher C.A. 2012. Macroevolutionary dynamics and historical biogeography of primate diversification inferred from a species supermatrix. *PLoS One.* 7:e49521.
- Swofford D.L. 2003. PAUP*: phylogenetic analysis using parsimony (*and other methods). Version 4. Sunderland (MA): Sinauer Associates.
- Swofford D.L., Waddell P.J., Huelsenbeck J.P., Foster P.G., Lewis P.O., Rogers J.S. 2001. Bias in phylogenetic estimation and its relevance to the choice between parsimony and likelihood methods. *Syst. Biol.* 50:525–539.
- Tarver J.E., dos Reis M., Mirarab S., Moran R.J., Parker S., O’Reilly J.E., King B.L., O’Connell M.J., Asher R.J., Warnow T., Peterson K.J., Donoghue P.C.J., Pisani D. 2016. The interrelationships of Placental Mammals and the Limits of Phylogenetic Inference. *Genome Biol. Evol.* 8:330–344.
- Thompson R.S., Bärmann E. V, Asher R.J. 2012. The interpretation of hidden support in combined data phylogenetics. *J. Zool. Syst. Evol. Res.* 50:251–263.
- 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.
- Wickham H., Averick M., Bryan J., Chang W., McGowan L.D., François R., Grolemund G., Hayes A., Henry L., Hester J., Kuhn M. 2019. Welcome to the Tidyverse. *J. Open Source Softw.* 4:1686.
- Wiens J.J. 2009. Paleontology, genomics, and combined-data phylogenetics: can molecular data improve phylogeny estimation for fossil taxa? *Syst. Biol.* 58:87–99.
- Wiens J.J., Kuczynski C.A., Townsend T., Reeder T.W., Mulcahy D.G., Sites Jr J.W. 2010. Combining phylogenomics and fossils in higher-level squamate reptile phylogeny: molecular data change the placement of fossil taxa. *Syst. Biol.* 59:674–688.
- Yonezawa T., Segawa T., Mori H., Campos P.F., Hongoh Y., Endo H., Akiyoshi A., Kohno N., Nishida S., Wu J., Jin H. 2017. Phylogenomics and morphology of extinct paleognaths reveal the origin and evolution of the ratites. *Curr. Biol.* 27:68–77.
- Zhou C.-F., Bhullar B.-A.S., Neander A.I., Martin T., Luo Z.-X. 2019. New Jurassic mammaliaform sheds light on early evolution of mammal-like hyoid bones. *Science.* 365:276–279.