scientific reports

OPEN

Statistical evaluation of character support reveals the instability of higher-level dinosaur phylogeny

David Černý✉ & Ashley L. Simonoff

The interrelationships of the three major dinosaur clades (Theropoda, Sauropodomorpha, and Ornithischia) have come under increased scrutiny following the recovery of conflicting phylogenies by a large new character matrix and its extensively modified revision. Here, we use tools derived from recent phylogenomic studies to investigate the strength and causes of this conflict. Using maximum likelihood as an overarching framework, we examine the global support for alternative hypotheses as well as the distribution of phylogenetic signal among individual characters in both the original and rescored dataset. We find the three possible ways of resolving the relationships among the main dinosaur lineages (Saurischia, Ornithischiformes, and Ornithoscelida) to be statistically indistinguishable and supported by nearly equal numbers of characters in both matrices. While the changes made to the revised matrix increased the mean phylogenetic signal of individual characters, this amplified rather than reduced their conflict, resulting in greater sensitivity to character removal or coding changes and little overall improvement in the ability to discriminate between alternative topologies. We conclude that early dinosaur relationships are unlikely to be resolved without fundamental changes to both the quality of available datasets and the techniques used to analyze them.

The relationships between the three major clades that comprise Dinosauria (Theropoda, Sauropodomorpha, and Ornithischia) have historically been uncertain. In his seminal work that coined the name Ornithischia for herbivorous dinosaurs characterized by an opisthopubic pelvis, Seeley[1] regarded this taxon as only distantly related to the theropods and sauropodomorphs, which he grouped together as Saurischia (Fig. 1). This view of dinosaur polyphyly later extended to Saurischia itself[2–6], leaving the ornithischians, theropods, and sauropodomorphs as three lineages independently descended from non-dinosaurian archosaurs of uncertain interrelationships.

The recognition of dinosaur monophyly failed to immediately clarify the relationships among the three clades, as the earliest arguments in favor of a monophyletic Dinosauria were coupled not only with the continued use of the Saurischia–Ornithischia dichotomy, but also with the seemingly contradictory suggestion that the ornithischians may have evolved from "prosauropods" (= early sauropodomorphs)[7–9]. In the mid-1980s, the latter scenario was formalized as a phylogenetic hypothesis linking Ornithischia and Sauropodomorpha to the exclusion of Theropoda[10–12] in a group variously termed Ornithischiformes[13] or Phytodinosauria[14] (Fig. 1). However, this hypothesis was abandoned after the first rigorous application of algorithmic phylogenetics to nonavian dinosaurs by Gauthier[15], who provided detailed character evidence uniting the theropods and sauropodomorphs into a monophyletic Saurischia, cementing a view of dinosaur phylogeny that would remain uncontested for the following three decades[16–18].

Using a large new dataset, Baron et al. ([19]; henceforth BEA) recently destabilized this view by lending support to the third possible (and previously unforeseen) way of resolving the branch in question, namely, a clade formed by Theropoda and Ornithischia to the exclusion of Sauropodomorpha, for which the authors adopted Huxley's[20] name Ornithoscelida (Fig. 1). Seven months later, Langer et al. ([21]; henceforth LEA) published a response using a rescored version of BEA's character matrix with 9 taxa added, which recovered the traditional Saurischia–Ornithischia dichotomy, albeit with virtually no statistical support. The resulting controversy surrounding early dinosaur phylogeny has not been resolved by subsequent attempts to further rescore or add supposed key taxa such as *Pisanosaurus*[22,23] and *Chilesaurus*[24–27], nor by the use of more sophisticated phylogenetic methods such as time-free Bayesian inference[28] and Bayesian tip-dating[29].

Despite the considerable interest generated by BEA's contribution, there have been few attempts to quantify the relative support for each of the candidate topologies, distinguish between low information content and internal

Department of the Geophysical Sciences, University of Chicago, 5734 South Ellis Avenue, Chicago, IL 60637, USA. ✉email: david.cerny1@gmail.com
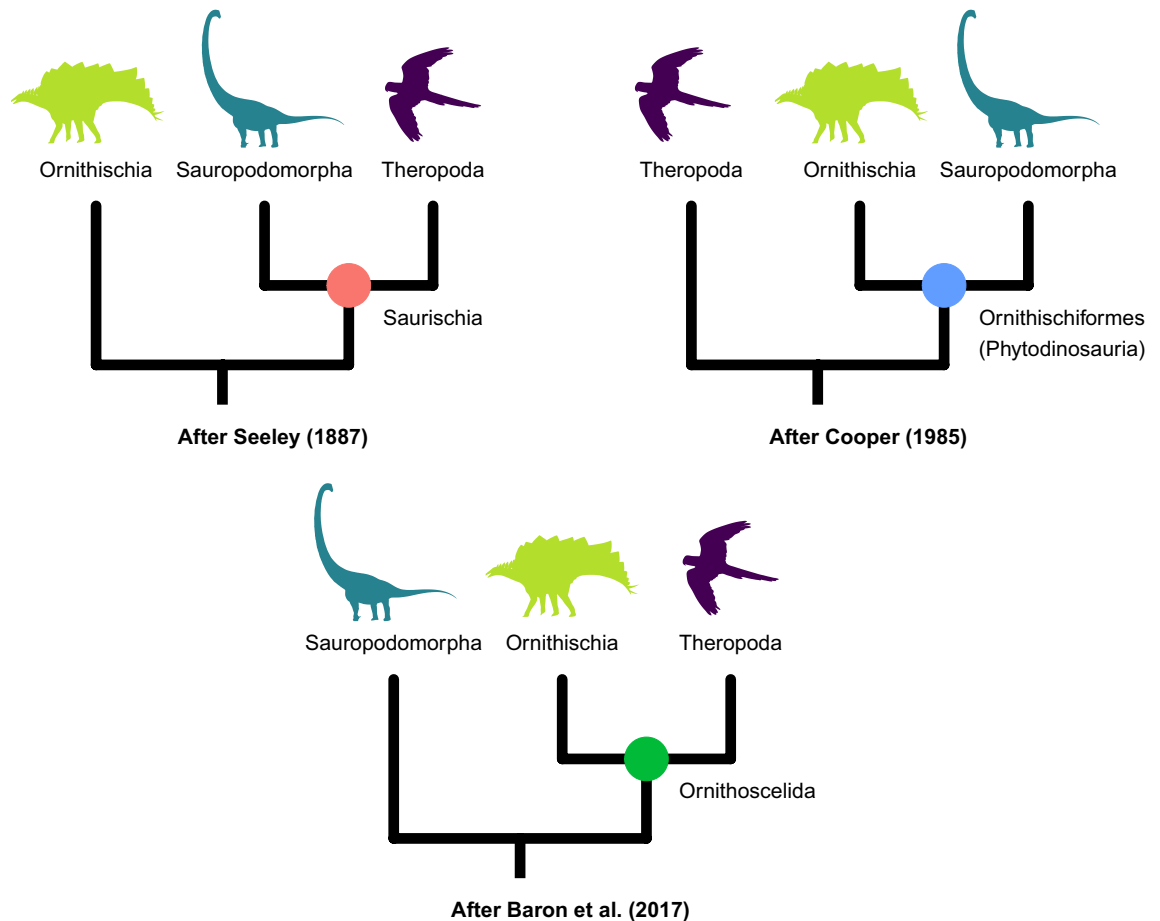
Figure 1. Three alternative hypotheses of early dinosaur phylogeny evaluated in this study. Dinosaur silhouettes from PhyloPic (http://phylopic.org) courtesy of Matthew Dempsey (CC BY 4.0), Jagged Fang Designs (CC0 1.0), and Sharon Wegner-Larsen (CC0 1.0). The figure was generated in R v4.2.2[35] using the packages *ape* v5.6-2[36,37], *ggtree* v3.6.2[38], *ggplot2* v3.4.0[39], *ggforce* v0.4.1[40], *rphylopic* v1.0.0[41], and *viridisLite* v0.4.1[42].

conflict, or identify the characters that may drive such conflict—lines of investigation that are more typical of phylogenomics[30–32] than morphological phylogenetics. LEA took early steps in this direction by demonstrating that when their data was analyzed in a parsimony framework, the reinstated Saurischia hypothesis was statistically indistinguishable from either of the alternatives[21]. However, subsequent studies have mostly reverted to reporting a single point estimate of the phylogeny, without testing whether its support significantly exceeded that of the next best hypothesis. The occasional attempts to use the number of extra steps relative to the most parsimonious tree for this purpose[19,24] suffer from the fact that this quantity has no statistical interpretation[33]. Moreover, despite the acknowledged centrality of character scoring differences to the conflict among the resulting topologies[21,22,25,26], only one study to date has attempted to determine which rescorings drove the difference between the hypotheses of early dinosaur phylogeny favored by BEA's and LEA's datasets[34], and its methodological scope was limited to parsimony.

Here, we use statistical tools drawn from phylogenomics to evaluate the relative support for the three hypotheses of large-scale dinosaur phylogeny in the BEA and LEA datasets, and to conduct detailed assessments of character support. We demonstrate the presence of pervasive conflict both across each dataset as a whole and among the subsets of characters with the strongest phylogenetic signal. We further show that although LEA's extensive changes to BEA's character coding dramatically altered the distribution of phylogenetic signal across the matrix, they did little to help discriminate among the three alternative topologies, which remain indistinguishable both before and after LEA's recoding. Our results suggest that there are many more plausible hypotheses of early dinosaur phylogeny than usually acknowledged, and that selecting between them may be beyond the reach of current character matrices and the techniques used to analyze them. We provide recommendations pertaining to both data-related and methodological aspects of the problem, and conclude that care should be taken to properly account for the uncertainty surrounding higher-level dinosaur phylogeny in downstream analyses.

Methods

Data. Our analyses were performed on the original character matrices used by BEA and LEA, whose properties are summarized in Table 1. Both were obtained from Graeme T. Lloyd's database of previously published character matrices with standardized formatting (http://graemetlloyd.com/matrdino.html; last accessed May 1,

	BEA	LEA
Taxa	74	83
Characters	457	457
Ordered	39	39
Constant	6	9
Binary	375	374
Of those autapomorphic:	45	43
Of those parsimony-informative:	330	331
Multistate	76	74
Of those autapomorphic:	2	1
Of those parsimony-informative:	74	73
Proportion missing	58.01%	55.58%

Table 1. Properties of the phylogenetic datasets analyzed in this study. The number of ordered characters is given as reported by BEA and LEA. For LEA, the proportion of missing data also includes the 109 polymorphic codings present in the original matrix, which are functionally equivalent to an unknown state under maximum likelihood.

2022). Both BEA and LEA used *Euparkeria capensis* and *Postosuchus kirkpatricki* as outgroup taxa. The two datasets differed in 3366 scorings (10.0% of the total number of overlapping cells) when polymorphic codings were treated as distinct from missing data, and in 3350 scorings (9.9% of the total number of overlapping cells) when treating the two as equivalent (as in our maximum likelihood analyses). Only 4 of the 74 overlapping taxa (*Dromomeron gigas*, *Dromomeron gregorii*, *Dromomeron romerii*, *Postosuchus kirkpatricki*) and 10 of the 457 characters (50–52, 118, 121–123, 152, 250, 344) were unaffected by the changes. The much higher number of differences previously reported in the literature (8050 scorings, or 21.2% of the total number of LEA's cells;[34]) also reflects non-overlapping cells (representing the extra taxa added by LEA) as well as the notational distinction between missing data ("?") and inapplicables ("-"), which has no analytical significance in currently used phylogenetic algorithms (but see[43–45]). Both datasets are organized by anatomical region, with cranial characters (1–146) followed by dental (147–185), axial (186–236), and pectoral (237–250) characters, and finally by characters pertaining to the forelimb (251–291), pelvis (292–351), and the hindlimb (352–457).

Maximum likelihood analyses. We used maximum likelihood (ML) in our analyses of BEA's and LEA's datasets for several reasons. First, we found it useful to explore how the resulting topologies might change when using a parametric rather than nonparametric approach to phylogenetic inference, since method-dependent results may indicate the presence of within-dataset conflict[46]. Second, despite its continued widespread use in the paleontological literature, maximum parsimony is well-known for its undesirable statistical properties compared to model-based methods[47], including in the context of morphological phylogenetics[48–51]. Third, topologically constrained ML analyses allowed us to directly compare support between the three hypotheses using a number of well-established and easily interpretable frequentist tests.

All maximum likelihood analyses were performed using IQ-TREE v2.1.3[52] with the default mix of starting trees (1 BioNJ + 99 parsimony trees). Both datasets were partitioned first into ordered and unordered characters and further by the number of character states, for a total of 6 partitions. In contrast to the number reported by the original authors (Table 1), only 36 and 37 characters from the BEA and LEA datasets were treated as ordered, respectively, since for characters 24, 334, and (for BEA) 180, only states 0 and 1 were observed. Branch lengths were treated as proportional among partitions (the -p command-line option in IQ-TREE), and unordered characters were assigned the M*k* model[53] with *k* ranging from 2 to 5. Constant characters (BEA: 29, 59, 150, 245, 248, 268; LEA: 29, 59, 75, 112, 139, 150, 248, 268, 288) were excluded, and an ascertainment bias correction[53] was applied to all partitions. Unlinked discrete gamma models of among-character rate heterogeneity were added to every substitution model except those applied to the character-poor unordered 5-state and ordered 4-state partitions.

Difficulty assessment and tree searches. A number of methods have been developed to quantify the expected difficulty of phylogenetic analysis or the amount of data needed to resolve a particular node[54–56]; however, these often rely on models that are inapplicable to morphological evolution, such as the multispecies coalescent[57]. To evaluate how challenging it would be to estimate early dinosaur interrelationships from the BEA and LEA datasets in a maximum-likelihood framework, we performed 100 topologically unconstrained tree searches on each dataset, and used the resulting ML trees to calculate a nonparametric difficulty measure recently proposed by Haag et al.[58]:

$$\text{difficulty} = \frac{1}{5}\left[\bar{d}_{\text{all}} \, + \, \bar{d}_{\text{pl}} \, + \, \frac{n'_{\text{all}}}{n_{\text{all}}} \, + \, \frac{n'_{\text{pl}}}{n_{\text{pl}}} \, + \, \left(1 - \frac{n_{\text{pl}}}{n_{\text{all}}}\right)\right], \tag{1}$$

where $\bar{d}_{\mathrm{all}}$ is the average pairwise normalized Robinson-Foulds (RF;[59]) distance between the $n_{\mathrm{all}} = 100$ inferred trees, n'_{all} is the number of unique topologies among the 100 ML trees, $\bar{d}_{\mathrm{pl}}$ is the average pairwise normalized RF distance within a subset of plausible trees, n_{pl} is the number of trees included in this set, and n'_{pl} is the number of unique topologies present in the plausible set. Each term of Eq. (1) ranges from 0 to 1, as does the overall difficulty score equal to their unweighted mean. Depending on the resulting value, ML phylogenetic inference can range from trivial (0) to effectively impossible (1). Following Morel et al.[60], the plausible tree set was constructed from all trees that were not found to be significantly worse than the best-scoring tree by any of the likelihood-based tests implemented in IQ-TREE. These were conducted using 10,000 approximate-bootstrap replicates (`-zb 10000 -zw -au`) generated by the resampling estimated log-likelihood method[61] and included the Kishino-Hasegawa test[62], the unweighted and weighted SH test[63], the approximately unbiased test[64], and expected likelihood weights[65]. Postprocessing was carried out in the R statistical computing environment[35] using the packages *phytools* v1.2-0[66], *TreeTools* v1.9.0[67], and their respective dependencies.

After the difficulty assessment, we performed one more round of unconstrained ML searches consisting of 10 runs (each with 100 starting trees). The overall ML estimate was obtained by selecting the best-scoring tree from the pooled sample of the 100 exploratory and 10 final runs. To assess clade support, we additionally performed ultrafast bootstrap approximation (UFBoot) with 1000 replicates (`-B 1000`), either after the fact (if the best-scoring tree was found during the exploratory runs) or simultaneously with the main tree search (for the last 10 runs). In addition to being relatively robust to model misspecification, the ultrafast bootstrap is less biased than the standard nonparametric bootstrap[68], and we also found it to be more numerically stable. Finally, we carried out topologically constrained analyses enforcing those hypotheses that were not supported by the unconstrained tree. Similar to the main analyses, each consisted of 10 runs performed simultaneously with 1000 UFBoot replicates.

Character-wise support. We followed the protocol of Shen et al.[31] to assess how support for the three competing hypotheses (S = Saurischia, Of = Ornithischiformes, Os = Ornithoscelida) was distributed across characters in both matrices, and to identify potential outliers. Using IQ-TREE, we estimated character-wise log-likelihood values for the ML trees yielded by both unconstrained and topologically constrained searches (`-wslr`). We first calculated the number of characters in either matrix that favored a given hypothesis (i.e., yielded the least negative log-likelihood value under it), and repeated this calculation for individual anatomical partitions. Multinomial tests were carried out using the R package *EMT* v1.2[69] to determine whether the resulting distributions were significantly different from uniform. We further determined the number of characters that ranked the three hypotheses identically in terms of their log-likelihoods between the two matrices. Next, we calculated the phylogenetic signal (PS) of the i-th character (C_i) in a given matrix following Eq. (5) of ref.[31]:

$$\mathrm{PS}_i = \frac{|\ln L(C_i \,|\, \mathrm{S}) - \ln L(C_i \,|\, \mathrm{Of})| \,+\, |\ln L(C_i \,|\, \mathrm{S}) - \ln L(C_i \,|\, \mathrm{Os})| \,+\, |\ln L(C_i \,|\, \mathrm{Of}) - \ln L(C_i \,|\, \mathrm{Os})|}{3} \quad (2)$$

Since the observed distributions of PS values were heavier-tailed than those shown in ref.[31], we used a different criterion to identify outliers, defining them as those characters whose PS was more than three standard deviations above the mean[70]. To assess the influence exerted by characters with strong phylogenetic signal on the resulting early dinosaur topologies, we generated 8 subsampled datasets by removing 1, 5, and 10 characters with the highest PS values from either matrix, as well as those characters whose PS values represented outliers according to the above criterion (BEA: 14 characters, LEA: 8 characters). These subsampled matrices were then subjected to unconstrained ML searches under the same settings as the original datasets (10 runs of 100 starting trees each + 1000 UFBoot replicates).

The definition of phylogenetic signal outlined above has recently been criticized for conflating cases in which one topology is strongly favored relative to either of the alternatives, and cases in which one topology is strongly disfavored relative to both alternatives that may nevertheless remain nearly indistinguishable from each other[70]. To tease apart these two scenarios, we further evaluated the difference in log-likelihood scores (ΔCLS) separately for each pair of hypotheses. For example, following Eq. (2) of ref.[31], the support afforded to Saurischia over Ornithoscelida by the i-th character was calculated as:

$$\Delta \mathrm{CLS}(\mathrm{S}, \mathrm{Os})_i = \ln L(C_i|\mathrm{S}) - \ln L(C_i|\mathrm{Os}) \quad (3)$$

We then applied the same criterion for outliers to absolute ΔCLS values derived from each pairwise comparison, and identified those characters that represented outliers in at least two of the three comparisons. These corresponded to characters that either strongly favored or strongly disfavored a given topology relative to both of the alternatives. We also applied the criterion suggested by Francis and Canfield[70] and identified those characters for which the log-likelihood difference between the best and second best hypotheses exceeded 0.5.

Rescoring of individual characters. To identify characters whose rescoring may have had disproportionate impact on the resulting topology, we ran further ML analyses on modified versions of both datasets in which we successively recoded one character at a time to its scoring in the opposite dataset. We excluded from consideration those characters whose coding did not change between the two matrices (see "Data") as well as those that would be rendered constant by reverting their scoring to that of the opposite dataset (see "Maximum likelihood analyses"), resulting in a total of 879 analyses. All of these were conducted under the same settings as the analyses of the original datasets (10 runs of 100 starting trees each + 1000 UFBoot replicates). Using a custom R script employing the package *phangorn* v2.10.0[71], we scored each resulting ML topology for the recovery of Saurischia, Ornithischiformes, or Ornithoscelida, and extracted the UFBoot value of whichever of these three

clades was present in the tree. To facilitate this process, we treated the names Saurischia, Ornithischiformes, and Ornithoscelida as referring to node-based clades, operationally defining them as (*Dilophosaurus + Plateosaurus*), (*Scelidosaurus + Plateosaurus*), and (*Dilophosaurus + Scelidosaurus*), respectively.

Results

Maximum likelihood analyses. Our initial round of 100 maximum likelihood (ML) searches suggested that estimating early dinosaur phylogeny from the two datasets (Table 1) would present considerable difficulty due to the ruggedness of the resulting tree space. The high difficulty scores (BEA = 0.596, LEA = 0.649) were mostly driven by a lack of topological congruence, since each of the 100 ML estimates had a unique topology that slightly (average pairwise normalized RF distance: BEA = 0.330, LEA = 0.395) but appreciably differed from those of the remaining trees (Supplementary Table 1). For the BEA dataset, the best-scoring tree was generally similar to the original parsimony estimate (Supplementary Fig. 1) and strongly supported a monophyletic Ornithoscelida (UFBoot = 99%). In contrast, the LEA dataset supported a version of the Ornithischiformes hypothesis which nested Ornithischia within Sauropodomorpha (Supplementary Fig. 2). The sister-group relationship between ornithischians and a clade of derived sauropodomorphs corresponding in content to Bagualosauria of ref.[72] again received substantial support (UFBoot = 97%), while the broader clade uniting Ornithischiformes proper with early sauropodomorphs was less robust (UFBoot = 94%).

Trees that were constrained to alternative early dinosaur topologies (Saurischia and Ornithischiformes for BEA, Saurischia and Ornithoscelida for LEA) exhibited log-likelihoods that fell well within the range yielded by the initial 100 unconstrained ML searches, suggesting that the three hypotheses were statistically indistinguishable for both datasets. We formally corroborated this result using multiple likelihood-based topology tests, all of which indicated that none of the three topologies was significantly better or worse than the others (Tables 2 and 3). The topological constraints were accommodated by the two datasets in markedly different ways. When their monophyly was enforced, Saurischia and Ornithischiformes were subtended by near-zero-length branches in the trees produced by the BEA dataset (Supplementary Figs. 3 and 4), indicating a lack of characters that could be convincingly interpreted as saurischian or ornithischiform synapomorphies. For the LEA dataset, enforcing the monophyly of Saurischia yielded a topology that closely resembled the original parsimony estimate (cf. Supplementary Fig. 5 and Fig. 1 of ref.[21]), while the tree optimized under an Ornithoscelida constraint showed an idiosyncratic topology that nested Ornithischia deep within theropods (Supplementary Fig. 6), reminiscent of hypotheses recently proposed by Baron[23] to account for stratigraphic incongruence in ornithischian origins.

Character-wise support. In both datasets, the distribution of character support for each of the three hypotheses was indistinguishable from uniform, both across the matrix as a whole (multinomial test; BEA: $p = 0.063$, LEA: $p = 0.668$) and within most of the individual anatomical partitions (Fig. 2b,d). Despite the overall similarity of character support distributions between the two matrices, we found substantial differences at the level of individual characters. Only 70 out of the 447 overlapping non-constant characters ranked the three hypotheses in the same way in terms of their log-likelihoods, less than the one-sixth expected by chance. Similarly, only 145 characters favored the same hypothesis, a number that is also indistinguishable from the one-third expected by chance (multinomial test: $p = 0.726$). While the number of characters supporting Ornithoscelida decreased as a result of LEA's rescoring (from 174 to 147) and the number of characters supporting Saurischia marginally increased (from 141 to 143), even in the LEA dataset, more characters supported Ornithoscelida than Saurischia, although the difference was negligible (Fig. 2b,d).

Examining support for competing hypotheses in terms of simple preference (i.e., by recording which tree yields the highest log-likelihood score for a given character) fails to account for the fact that most characters do not strongly prefer any of the three alternatives, rendering the resulting log-likelihood differences negligible and overly sensitive to minor branch length differences. Indeed, the character-wise log-likelihood difference (ΔCLS) between the best and second best hypotheses exceeds a threshold of 0.5 for less than 10% of characters in the BEA dataset and 30% of characters in the LEA dataset, although we note that these proportions are still substantially higher than is typical of the phylogenomic datasets for which the threshold was originally defined[31,70]. In the BEA dataset, the majority of these "strong" characters (29 out of 45) support Ornithoscelida, which is also the case for a plurality (63 out of 137) of such characters in the LEA dataset.

To obtain a more fine-grained view of key characters driving the support for particular topologies, we employed the method suggested by Shen et al.[31] and calculated the phylogenetic signal (PS) of each character as

Hypothesis	ln L	bp (RELL)	p (KH)	p (SH)	p (wSH)	p (AU)	c (ELW)
Ornithoscelida	− 6368.110126	0.754 +	0.826 +	1 +	0.946 +	0.824 +	0.748 +
Saurischia	− 6377.249639	0.162 +	0.174 +	0.26 +	0.254 +	0.218 +	0.163 +
Ornithischiformes	− 6377.484497	0.0844 +	0.114 +	0.247 +	0.194 +	0.222 +	0.0888 +

Table 2. Relative fit of the three alternative early dinosaur topologies to the BEA dataset. Best (unconstrained) and constrained ML trees are listed in order of decreasing likelihood (*L*). The bootstrap proportion (bp) from the resampling estimated log-likelihood (RELL) method is shown along with the *p*-values from the Kishino-Hasegawa (KH), Shimodaira-Hasegawa (SH), weighted Shimodaira-Hasegawa (wSH), and approximately unbiased (AU) tests, as well as the confidence value (*c*) based on expected likelihood weight (ELW). Significant rejection of (−) or the failure to reject (+) a given topology is indicated next to each test statistic.
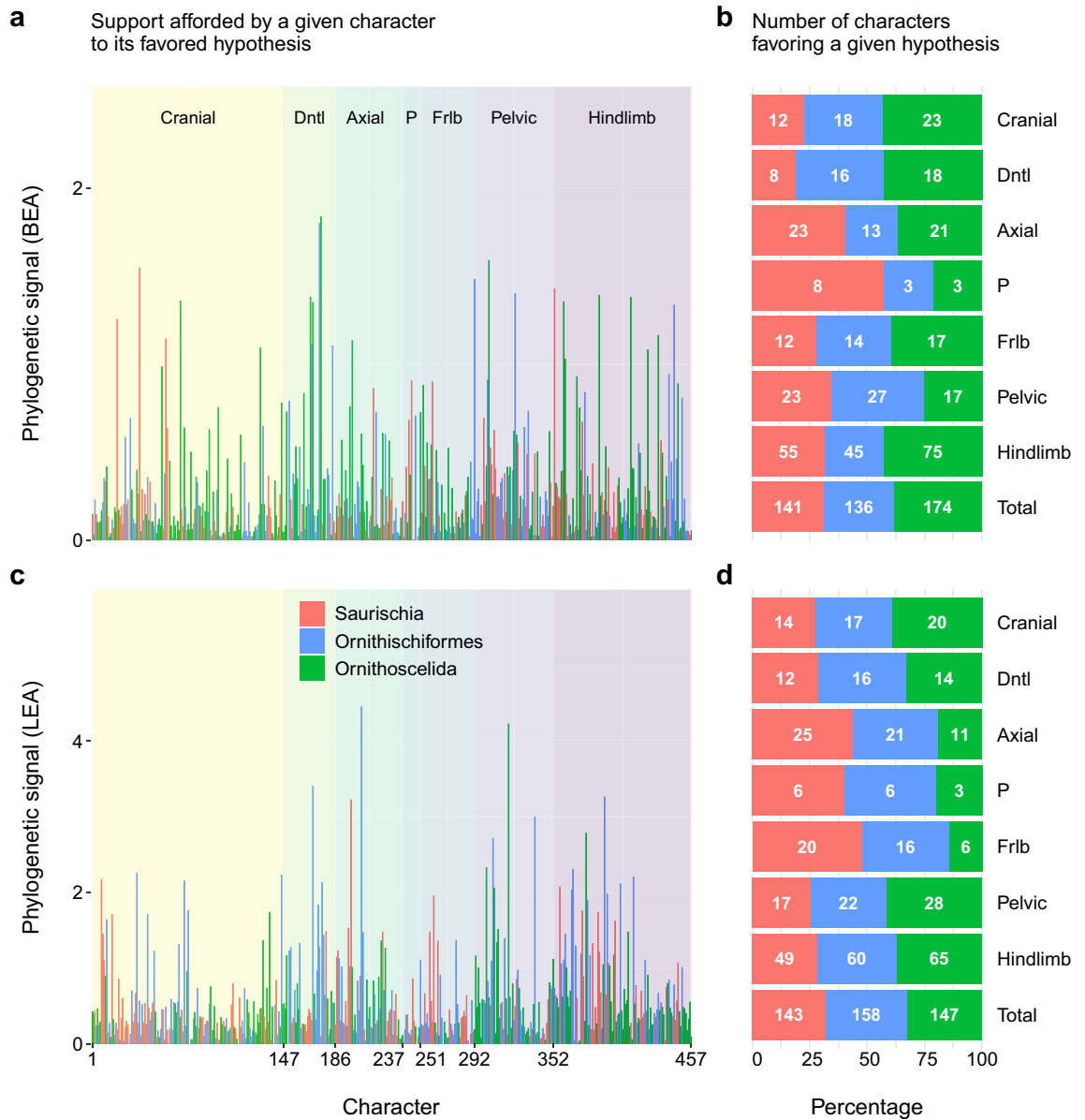
Figure 2. Distribution of phylogenetic signal across the two datasets. Phylogenetic signal of every non-constant character from the BEA (**a**) and LEA (**c**) dataset is plotted against its placement in the matrix and colored by its favored (highest-likelihood) hypothesis. The numbers and proportions of characters favoring a given hypothesis are further shown for the BEA (**b**) and LEA (**d**) datasets as well as the individual anatomical regions by which they are organized. Abbreviations: Dntl = dental, P = pectoral, Frlb = forelimb. The plots were generated in R v4.2.2[35] using the packages *scales* v1.2.1[73], *ggplot2* v3.4.0[39], *cowplot* v1.1.1[74], *patchwork* v1.1.2[75], *ggnewscale* v0.4.8[76], and *viridisLite* v0.4.1[42].

Hypothesis	ln *L*	bp (RELL)	*p* (KH)	*p* (SH)	*p* (wSH)	*p* (AU)	*c* (ELW)
Ornithischiformes	− 7023.11918	0.723 +	0.882 +	1 +	0.962 +	0.832 +	0.718 +
Saurischia	− 7034.61052	0.102 +	0.118 +	0.29 +	0.195 +	0.234 +	0.106 +
Ornithoscelida	− 7035.848825	0.175 +	0.181 +	0.273 +	0.246 +	0.208 +	0.177 +

Table 3. Relative fit of the three alternative early dinosaur topologies to the LEA dataset. Best (unconstrained) and constrained ML trees are listed in order of decreasing likelihood. Abbreviations as in Table 2.

the mean of the absolute values of the three pairwise log-likelihood differences (Saurischia vs. Ornithischiformes, Saurischia vs. Ornithoscelida, Ornithischiformes vs. Ornithoscelida). Relative to the BEA dataset, the LEA dataset displays substantially higher mean (0.612 vs. 0.295) as well as maximum (4.454 vs. 1.840) phylogenetic signal values, consistent with its higher number of strong characters. The PS values of individual characters also differ considerably between the two datasets (Fig. 2a,c). In particular, among the characters with outlier PS values (more than three standard deviations above the mean), only one (character 169, serrations of maxillary and dentary teeth) is shared between the BEA matrix (in descending order of PS: 175, 174, 303, 37, 292, 353, 323, 387, 167, 411, 68, 360, 169, 444) and the LEA matrix (206, 318, 169, 391, 198, 338, 377, 306). Gradual removal of 1, 5, or 10 characters with the highest PS values or of all characters whose PS values represented outliers never caused either matrix to switch from its preferred topology to an alternative one (Supplementary Figs. 7–14). The exclusion of high-PS characters made little difference to the high statistical support for Ornithoscelida in the BEA dataset (UFBoot = 96–99%; Supplementary Figs. 7–10) but caused a drastic erosion of support for Ornithischiformes in the LEA dataset (Ornithischia + Bagualosauria: UFBoot = 69–93%; Ornithischiformes proper + early sauropodomorphs: UFBoot = 52–79%; Supplementary Figs. 11–14).

To further distinguish among different ways in which a character can attain a high PS value, we separately compared each pair of hypotheses (Fig. 3), focusing on those characters that also emerged as outliers in at least two of the three pairwise comparisons. These corresponded to characters that strongly favored a particular hypothesis, strongly disfavored one, or both. The results reveal substantial conflict among the high-PS characters, as each matrix contains characters that both strongly favor and strongly disfavor its globally preferred topology (Ornithoscelida for BEA, Ornithischiformes for LEA) as well as one or both of its alternatives (Fig. 3, Table 4). Neither of the characters that strongly favor Ornithoscelida in the BEA dataset was among the 21 synapomorphies of this clade originally identified by BEA, although one such synapomorphy (character 360, state 1: medial bowing of the femur forming a gentle curve) was found to strongly disfavor Saurischia in the present analysis (Table 4). Similarly, there is virtually no overlap between the characters strongly favoring or disfavoring one of the three hypotheses in the present study, and the "keystone" characters recently identified using a parsimony-based approach[34].

Rescoring of individual characters. To determine which of LEA's coding changes most contributed to the difference between the topologies yielded by the original and rescored datasets, we successively replaced the scoring of each character in either matrix by its scoring from the opposite matrix, and checked whether this
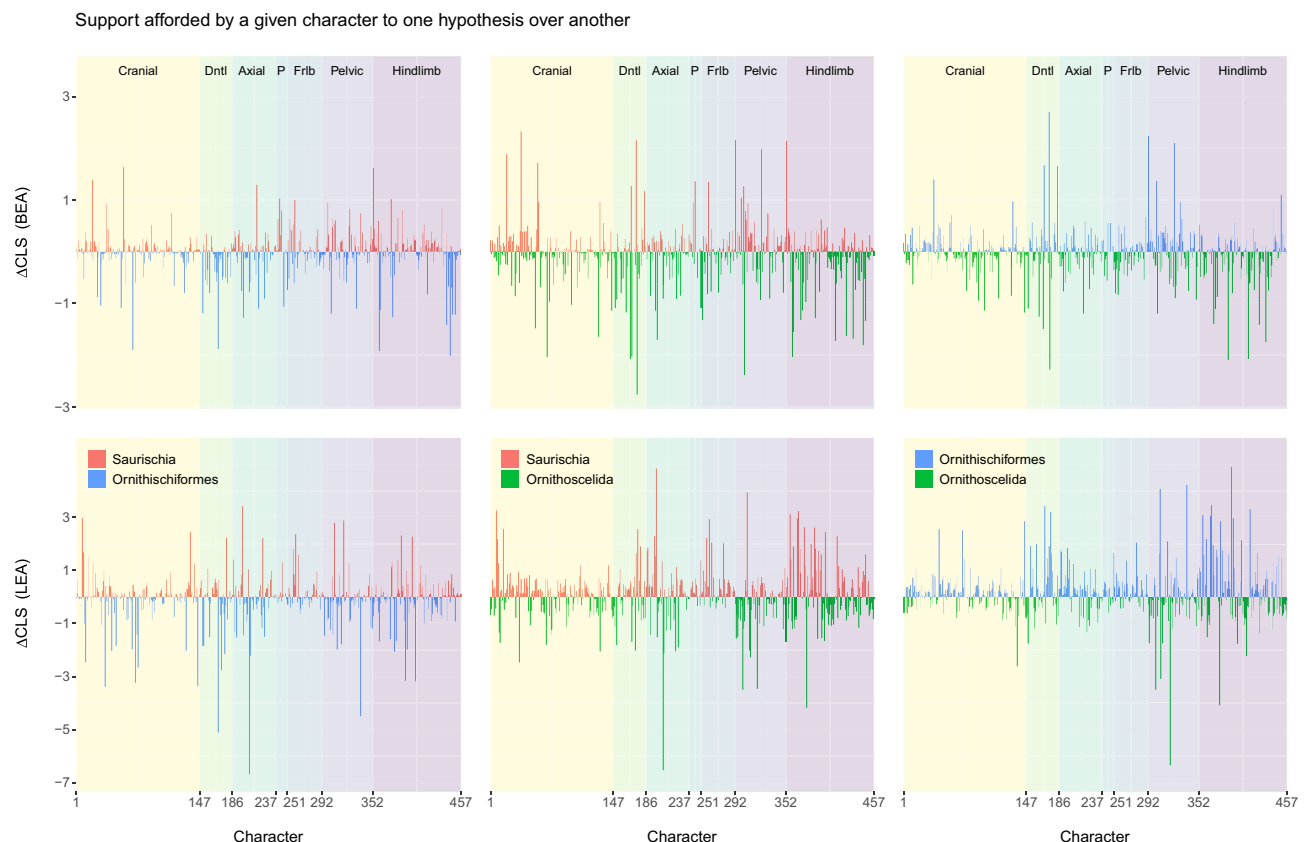
Support afforded by a given character to one hypothesis over another

Figure 3. Pairwise comparisons of character support for the three alternative early dinosaur topologies. For each pair of hypotheses, the difference in character-wise log-likelihood values (ΔCLS) of every non-constant character from the BEA (top) and LEA (bottom) datasets is plotted against its placement in the matrix. Anatomical region abbreviations as in Fig. 2. The plots were generated in R v4.2.2[35] using the packages *scales* v1.2.1[73], *ggplot2* v3.4.0[39], *cowplot* v1.1.1[74], *patchwork* v1.1.2[75], *ggnewscale* v0.4.8[76], and *viridisLite* v0.4.1[42].

change was sufficient for the re-estimated ML tree to switch to a different topology (Fig. 4). The procedure had little impact on the BEA dataset's support for Ornithoscelida, which proved robust to the rescoring of any one of the 438 applicable characters and remained high on average (mean UFBoot = 98.8%; Fig. 4a). Only three characters (110, 114, 387) caused the support for Ornithoscelida to drop below the 95% threshold when changed to their scorings in the LEA dataset. Of these, none was optimized as an ornithoscelidan synapomorphy in BEA's original analysis[19], and while all favored Ornithoscelida over the alternatives in the original BEA matrix, only character 387 did so strongly (ΔCLS > 0.5 relative to the next best hypothesis). In the LEA matrix, all three characters ranked Saurischia and Ornithoscelida as the best- and worst-supported hypothesis, respectively, but the log-likelihood difference between the two exceeded the 0.5 threshold only for character 387.

In contrast, reverting four characters to their original scoring in the BEA matrix proved sufficient to flip the ML result from the LEA dataset (Ornithischiformes) either to Ornithoscelida (characters 148, 363, 370) or to Saurischia (character 77) (Fig. 4b), producing highly idiosyncratic topologies (Supplementary Figs. 15–18). In the ornithoscelidan trees, Ornithischia was deeply nested within theropods (with *Panguraptor* and *Zupaysaurus* consistently recovered closer to Ornithischia than to other theropods), similar to the results obtained from a constrained analysis of the unmodified LEA matrix (Supplementary Fig. 6). The support for the nodes uniting Ornithischia with its successive theropod outgroups was generally low, although one such node received a UFBoot value of 97% in the analysis based on the rescoring of character 370 (Supplementary Fig. 18), consistent with the fact that state 2 of this character (prominent, wing-like anterior trochanter) represented an ornithoscelidan synapomorphy under BEA's original scoring ([19]; see also[24]). The only analysis to recover a monophyletic Saurischia did so with poor support, and found unconventional relationships elsewhere in the tree (e.g., ornithischian herrerasaurids; Supplementary Fig. 15). None of the characters that caused a switch from Ornithischiformes to Ornithoscelida favored Ornithischiformes in the LEA dataset, and only two of them (characters 148 and 370) favored Ornithoscelida in the BEA dataset. Notably, character 77 consistently ranked Saurischia as the worst of the three hypotheses under both BEA and LEA scorings; the fact that its rescoring was nevertheless sufficient for Saurischia to emerge as the preferred hypothesis indicates an extreme degree of instability across the LEA matrix. This is also borne out by the fact that even among those trees which continued to support Ornithischiformes, altering the scoring of a single character caused the UFBoot support for this clade to drop from 97% to (on average) 89.5% (Fig. 4b). Only 105 of the 441 applicable characters (23.8%) upheld Ornithischiformes with UFBoot support greater than 95% upon rescoring.

Discussion

By repurposing a protocol originally developed for phylogenomic data[31], we found that both the BEA and LEA datasets are unable to conclusively resolve the interrelationships of major dinosaur clades. All three hypotheses of overall dinosaur phylogeny—Saurischia, Ornithischiformes, and Ornithoscelida—remain plausible, and neither dataset shows any of these to be significantly better or worse than the alternatives (Tables 2 and 3). Our results suggest that this is not due to low information content of the two matrices; in fact, the proportion of characters that strongly discriminate between the best and second best hypotheses (ΔCLS > 0.5) is far higher (10–30%) than typical for phylogenomic data (0.1–7%;[31,70]). Similarly, although both matrices exhibit a highly uneven

Character	PS in BEA	PS in LEA	Outlier in BEA	Outlier in LEA
68	1.360	0.352	S (−)	−
167	1.383	0.409	Os (+)	−
169	1.352	3.407	S (−)	Of (+)
174	1.801	1.275	Os (−)	−
175	1.840	0.103	Os (+)	−
198	0.139	3.228	−	S (+)
206	0.605	4.454	−	S (−)
292	1.483	0.193	Os (−)	−
306	0.262	2.716	−	Os (−)
318	0.370	4.231	−	Os (+), Of (−)
323	1.403	0.500	Os (−)	−
338	0.493	2.998	−	Of (+)
353	1.429	0.738	S (+)	−
360	1.356	1.112	S (−)	−
377	0.057	2.787	−	Os (+)
391	0.226	3.267	−	Of (+)
444	1.337	0.780	S (−)	−

Table 4. Outlier characters in the BEA and LEA matrices. Of = Ornithischiformes, Os = Ornithoscelida, S = Saurischia. Characters were only included if both their phylogenetic signal (PS) and at least two of the three pairwise log-likelihood differences [ΔCLS(S, Of), ΔCLS(S, Os), ΔCLS(Of, Os)] were more than three standard deviations above the mean. Hypotheses that are strongly favored (+) or disfavored (−) relative to both alternatives are indicated next to each character.
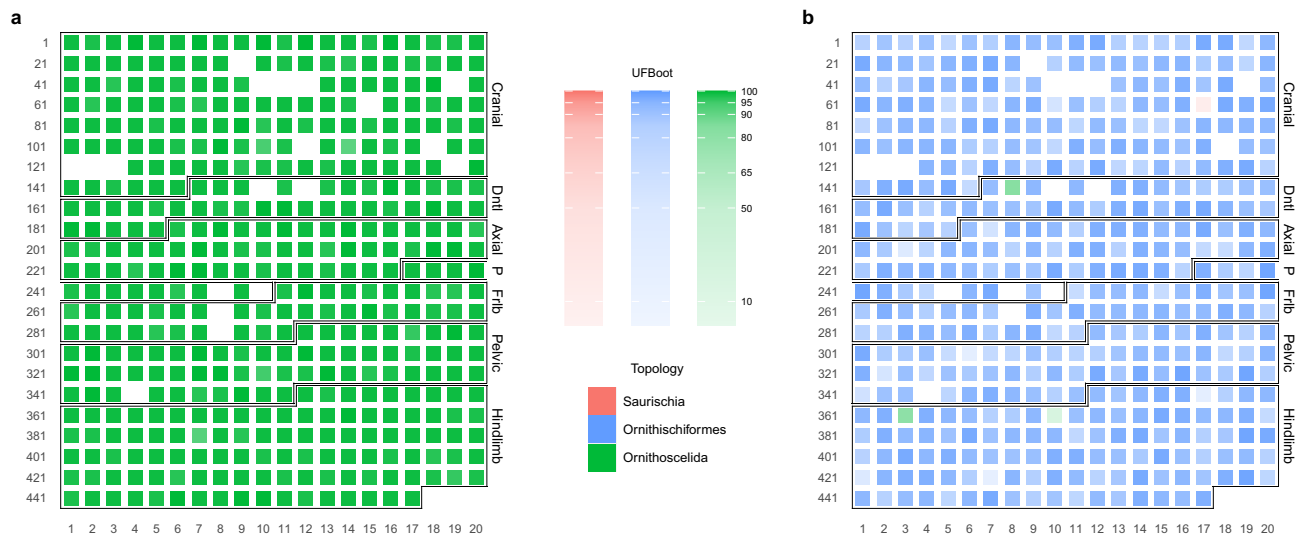
Figure 4. Support for alternative topologies after rescoring each matrix one character at a time. Characters changed from BEA's to LEA's coding (**a**) or from LEA's to BEA's coding (**b**) are arranged by row following their placement in either dataset; black boxes denote different anatomical regions. Color and opacity of individual heatmap cells denote the topology of the ML tree inferred after rescoring the character in question and the ultrafast bootstrap (UFBoot) support for that topology's focal clade, respectively. Gaps correspond to characters whose coding did not differ between the two matrices or characters that would be rendered constant by having their coding changed to that of the opposite matrix. Anatomical region abbreviations as in Fig. 2. The plots were generated in R v4.2.2[35] using the packages *ape* v5.6-2[36,37], *dplyr* v1.0.10[77], *phangorn* v2.10.0[71], *scales* v1.2.1[73], *ggplot2* v3.4.0[39], *cowplot* v1.1.1[74], *patchwork* v1.1.2[75], and *ggnewscale* v0.4.8[76].

distribution of phylogenetic signal and contain several outlier characters strongly favoring or disfavoring particular topologies, the results were not exclusively driven by a handful of such outliers, since their removal had limited impact on the support for one topology or another (Supplementary Figs. 7–14). Instead, we hypothesize that the lack of meaningful statistical support for any of the three hypotheses (partially obscured by high UFBoot values) is due to pervasive conflict among individual characters. Limiting the focus to the characters with the highest PS values revealed patterns of conflict (Table 4) that were similar to those observed across each matrix as a whole (Fig. 2), explaining why their removal did little to change the underlying distribution of support.

According to almost every metric employed in this study, the LEA dataset produces less stable results than the BEA dataset. It yields a higher phylogenetic difficulty score (Supplementary Table 1), exhibits a more uniform distribution of character support for the three hypotheses (Fig. 2c,d), and is more sensitive to the exclusion of outlier characters (Supplementary Figs. 11–14) despite containing fewer of them (Table 4). The LEA matrix was also less robust to changes in character coding, as demonstrated by its tendency to flip the ML tree from one topology to another after reverting a single character to its original scoring by BEA (Fig. 4). Consistent with the findings above, this greater degree of instability was not caused by weaker phylogenetic signal in the LEA matrix. In fact, although LEA's rescorings and taxon additions only marginally improved the completeness of the matrix (Table 1), they resulted in a markedly higher mean character-wise phylogenetic signal (Supplementary Fig. 19), and amplified the log-likelihood differences between competing hypotheses both for the dataset as a whole (Table 3) and for individual characters (Supplementary Fig. 20). In effect, the stronger phylogenetic signal present in the recoded and expanded matrix only served to amplify, rather than eliminate, underlying conflict within the dataset. In light of the failure to resolve this pervasive conflict by extensive coding changes, we outline several alternative recommendations for identifying its sources and assessing the relative support for competing topologies in its presence.

First, we recommend re-examining the original dataset at a deeper level, as the recovery of divergent yet statistically indistinguishable topologies may serve to highlight fundamental issues in the underlying character data. Poorly formulated characters should ideally be redefined rather than simply rescored. For example, LEA's coding changes that caused character 174 (recurvature of maxillary and dentary teeth) to lose the outlier status it originally had in the BEA dataset (Table 4) still took place in the framework of the vague definition inherited from BEA, who in turn modified it from an even earlier study[78]. The character description provides no quantitative criterion for differentiating between teeth possessing strong, weak, or no recurvature, allowing for more or less arbitrary coding changes. Indeed, the scoring of *Efraasia* was changed by LEA from no recurvature to weak recurvature without explicit justification or photographic evidence, and on the basis of the same published sources which BEA cited in support of their own original coding. The problem of vague character descriptions and subjective scoring decisions is widespread in the two matrices (e.g., characters 114, 216, 266, 337) and compounded by a number of additional issues, some of which were noted by BEA and LEA themselves. These include multiple instances of scoring taxa for characters that cannot be ascertained from their known material[22]. While this practice may occasionally be justifiable (e.g., using the length of the mandible to estimate the length

of the skull), coding taxa based on assumed rather than observed morphologies represents a potentially serious biasing factor[79,80].

Second, some of the issues previously suggested to represent fundamental problems with the data may in fact be mitigated at the methodological level. For example, character non-independence[21] violates the assumptions of the methods currently used to analyze paleontological datasets (including those employed in this study), but frameworks capable of dealing with hierarchical and correlated characters are under development[45], as are ontology-based methods for their semi-automated detection and characterization[81,82]. More advanced methods may ultimately also alleviate other problems with existing matrices. The BEA and LEA datasets contain instances of problematic state delimitation that make it impossible to assign certain morphologies to any existing state (e.g., neither of the two states defined for character 28 can account for an antorbital fenestra equal in size to 10–15% of skull length), or merge distinct morphologies into a single state (e.g., character 77, state 0: paraquadratic foramen small or absent). Both exemplify a more general problem, namely the arbitrary discretization of characters that would be more naturally treated as continuous[83]. This ubiquitous feature of morphological phylogenetic datasets reflects long-standing methodological limitations, which have only recently been overcome by phylogenetic software packages that simultaneously implement models for discrete and continuous characters, making it possible to combine both types of data in a single analysis[84,85].

Third, instead of attempting to rescore or redefine an entire dataset in an indiscriminate "shotgun" approach, it is prudent to determine which characters are responsible for the signal in that dataset's results. While LEA re-examined the putative ornithoscelidan synapomorphies identified by BEA at an admirable level of detail[21], both previous findings[34] and our own results demonstrate that the characters supporting a particular topology do not always coincide with the characters that map as the synapomorphies of that topology's focal clade. Methods for identifying such critical characters are now available in parsimony[34], maximum-likelihood[31,70], and Bayesian[82] frameworks, and can be profitably used to narrow the focus of potential rescoring efforts, which often involve the time- and labor-intensive process of gathering data from first-hand observations of multiple museum specimens. The benefits of comprehensively revising a pre-existing character matrix have to be weighed against the costs inherent to such an effort, as well as the risk of introducing new errors into the data. In morphological phylogenetics, extensive reuse and iterative expansion of pre-existing matrices gives rise to complicated dataset genealogies[80,86] in which coding error can propagate and compound over time. On the other hand, our results suggest that the effort invested into the comprehensive rescoring of a large pre-existing dataset can be difficult to justify: the three hypotheses of early dinosaur phylogeny were statistically indistinguishable based on the original BEA dataset, and remain such after LEA's extensive revision of it.

Fourth, we urge paleontologists to quantify the uncertainty associated with their phylogenetic hypotheses using well-characterized tools with a clear statistical interpretation. When alternative ways of resolving a given branch are of interest, as in the controversy surrounding early dinosaur phylogeny, we encourage the community to move beyond the mere reporting of a phylogenetic point estimate toward explicitly testing it against the next best alternative. Following recent practice[87], we used a variety of likelihood ratio tests (LRTs) to this end[63–65], but other approaches are possible. Bayesian inference differs from maximum likelihood in its treatment of nuisance parameters such as branch lengths or the parameters of the substitution and rate heterogeneity models, which are jointly optimized with the parameter of interest (topology) in maximum likelihood but marginalized over in Bayesian methods[88]. Both approaches have their advantages[89], and as a result, Bayes factors—a Bayesian equivalent of LRTs, relying on marginal rather than joint likelihoods[90]—can represent a useful alternative way of evaluating competing topologies[91,92]. Although most of the best-performing marginal likelihood estimators are much more computationally demanding than joint likelihood inference[93], the minute size of phylogenetic datasets employed by paleontologists makes their application relatively easy, and Bayes factor topological comparisons are accordingly starting to see use in dinosaur phylogenetics[94].

Taken together, our results suggest that large-scale dinosaur phylogeny is much more poorly understood than commonly acknowledged. In particular, despite the widespread perception that saurischian monophyly is challenged primarily by the recently proposed Ornithoscelida hypothesis[95–97], the earlier Ornithischiformes hypothesis receives comparable support, and is in fact weakly preferred when the LEA dataset is analyzed using maximum likelihood (Table 3). Additional support for Ornithischiformes was also recently detected in an independent dataset[98]. Moreover, not only the three major hypotheses—Saurischia, Ornithischiformes, and Ornithoscelida—but also a number of their variations nesting the ornithischians deep within Sauropodomorpha or Theropoda (Supplementary Figs. 2, 6, 16–18) cannot be ruled out at present. Indeed, the specific variation on the Ornithischiformes topology recovered in this study shows the ornithischians to be not just sister to, but rather nested within Sauropodomorpha, with a clade of Carnian to early Norian sauropodomorphs (approximately corresponding to Guaibasauridae of ref.[99] or Saturnaliidae of ref.[72]) branching off before the ornithischians. This result is consistent with a phylogeny inferred from the LEA dataset by Parry et al.[28] using time-free Bayesian inference, showing that the two model-based methods yield topologies that are much more similar to each other than to those favored by parsimony. By positing an early divergence of the early Late Triassic sauropodomorphs, this scenario helps reduce the temporal gap between the first appearance of Ornithischia and its sister group[23], and shows remarkable congruence with the early suggestions that the ornithischians may have arisen from within "prosauropods"[7–9,14].

Our findings suggest that higher-level dinosaur interrelationships represent a phylogenetic problem of considerable difficulty that is unlikely to be conclusively resolved by minor additions and superficial modifications to the datasets currently in use. While our investigation was limited to two such datasets[19,21], there are few reasons to believe that other matrices currently employed by dinosaur paleontologists are free of the problems identified here. Indeed, the main alternative to the datasets examined here has repeatedly lent support to yet another nonstandard hypothesis nesting the putatively non-dinosaurian Silesauridae within Ornithischia[96,100,101], indicating that the number of plausible early dinosaur phylogenies proliferates even further when not only

the three major clades, but also species-poor lineages such as Herrerasauridae and Silesauridae are taken into consideration. As a result, the increasingly common practice of repeating comparative analyses under the Saurischia and Ornithoscelida topologies[95,97,102] most likely severely understates the uncertainty associated with early dinosaur phylogeny, potentially leading to biased or overconfident conclusions.

Data availability

All data matrices, tree files, log files, R code, and shell scripts are available from Zenodo: (https://doi.org/10.5281/zenodo.7979718).

References

1. Seeley, H. G. I. On the classification of the fossil animals commonly named Dinosauria. *Proc. R. Soc. Lond.* **43**, 165–171 (1887).
2. Romer, A. S. *Osteology of the Reptiles* (University of Chicago Press, 1956).
3. Charig, A. J., Attridge, J. & Crompton, A. W. On the origin of the sauropods and the classification of the Saurischia. *Proc. Linn. Soc. Lond.* **176**, 197–221 (1965).
4. Reig, O. A. The Proterosuchia and the early evolution of the archosaurs; An essay about the origin of a major taxon. *Bull. Mus. Comp. Zool.* **139**, 229–292 (1970).
5. Thulborn, R. A. Dinosaur polyphyly and the classification of archosaurs and birds. *Aust. J. Zool.* **23**, 249–270 (1975).
6. Charig, A. J. Problems in dinosaur phylogeny: A reasoned approach to their attempted resolution. *Geobios* **15**, 113–126 (1982).
7. Bakker, R. T. & Galton, P. M. Dinosaur monophyly and a new class of vertebrates. *Nature* **248**, 168–172 (1974).
8. Bonaparte, J. F. *Pisanosaurus mertii* Casamiquela and the origin of the Ornithischia. *J. Paleontol.* **50**, 808–820 (1976).
9. Cooper, M. R. The prosauropod dinosaur *Massospondylus carinatus* Owen from Zimbabwe: Its biology, mode of life and phylogenetic significance. *Occ. Pap. Natl. Mus. Monu. Rhod. B Nat. sci.* **6**, 689–840 (1981).
10. Paul, G. S. The segnosaurian dinosaurs: Relics of the prosauropod-ornithischian transition?. *J. Vertebr. Paleontol.* **4**, 507–515 (1984).
11. Paul, G. S. *Predatory Dinosaurs of the World: A Complete Illustrated Guide* (Simon and Schuster, 1988).
12. Sereno, P. C. The phylogeny of the Ornithischia: A reappraisal. In *Third Symposium on Mesozoic Terrestrial Ecosystems, Short Papers* (eds Reif, W.-E. & Westphal, F.) 219–226 (Attempto Verlag, 1984).
13. Cooper, M. R. A revision of the ornithischian dinosaur *Kangnasaurus coetzeei* Haughton, with a classification of the Ornithischia. *Ann. S. Afr. Mus.* **95**, 281–317 (1985).
14. Bakker, R. T. *The Dinosaur Heresies: New Theories Unlocking the Mystery of the Dinosaurs and Their Extinction* (William Morrow & Co., 1986).
15. Gauthier, J. A. Saurischian monophyly and the origin of birds. *Mem. Calif. Acad. Sci.* **8**, 1–55 (1986).
16. Novas, F. E. Dinosaur monophyly. *J. Vertebr. Paleontol.* **16**, 723–741 (1996).
17. Benton, M. J. Origin and relationships of Dinosauria. In *The Dinosauria* 2nd edn (eds Weishampel, D. B. *et al.*) 7–19 (University of California Press, 2004).
18. Nesbitt, S. J. The early evolution of archosaurs: Relationships and the origin of major clades. *Bull. Am. Mus. Nat. Hist.* **352**, 1–292 (2011).
19. Baron, M. G., Norman, D. B. & Barrett, P. M. A new hypothesis of dinosaur relationships and early dinosaur evolution. *Nature* **543**, 501–506 (2017).
20. Huxley, T. H. On the classification of the Dinosauria, with observations on the Dinosauria of the Trias. *Q. J. Geol. Soc.* **26**, 32–51 (1870).
21. Langer, M. C. *et al.* Untangling the dinosaur family tree. *Nature* **551**, E1–E3 (2017).
22. Baron, M. G. *et al.* Reply. *Nature* **551**, E4–E5 (2017).
23. Baron, M. G. *Pisanosaurus mertii* and the Triassic ornithischian crisis: Could phylogeny offer a solution?. *Hist. Biol.* **31**, 967–981 (2019).
24. Baron, M. G. & Barrett, P. M. A dinosaur missing-link? *Chilesaurus* and the early evolution of ornithischian dinosaurs. *Biol. Lett.* **13**, 20170220 (2017).
25. Müller, R. T., Pretto, F. A., Kerber, L., Silva-Neves, E. & Dias-da Silva, S. Comment on 'A dinosaur missing-link? *Chilesaurus* and the early evolution of ornithischian dinosaurs'. *Biol. Lett.* **14**, 20170581 (2018).
26. Baron, M. G. & Barrett, P. M. Support for the placement of *Chilesaurus* within Ornithischia: A reply to Müller *et al.*. *Biol. Lett.* **14**, 20180002 (2018).
27. Müller, R. T. & Dias-da Silva, S. Taxon sample and character coding deeply impact unstable branches in phylogenetic trees of dinosaurs. *Hist. Biol.* **31**, 1089–1092 (2019).
28. Parry, L. A., Baron, M. G. & Vinther, J. Multiple optimality criteria support Ornithoscelida. *R. Soc. Open Sci.* **4**, 170833 (2017).
29. Griffin, C. T. *et al.* Africa's oldest dinosaurs reveal early suppression of dinosaur distribution. *Nature* **609**, 313–319 (2022).
30. Reddy, S. *et al.* 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 (2017).
31. Shen, X.-X., Hittinger, C. T. & Rokas, A. Contentious relationships in phylogenomic studies can be driven by a handful of genes. *Nat. Ecol. Evol.* **1**, 0126 (2017).
32. Pease, J. B., Brown, J. W., Walker, J. F., Hinchliff, C. E. & Smith, S. A. Quartet sampling distinguishes lack of support from conflicting support in the green plant tree of life. *Am. J. Bot.* **105**, 385–403 (2018).
33. Felsenstein, J. *Inferring Phylogenies* (Sinauer Associates, 2004).
34. Goloboff, P. A. & Sereno, P. C. Comparative cladistics: Identifying the sources for differing phylogenetic results between competing morphology-based datasets. *J. Syst. Palaeontol.* **19**, 761–786 (2021).
35. R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria (2021). http://www.R-project.org/.
36. Paradis, E., Claude, J. & Strimmer, K. APE: Analyses of phylogenetics and evolution in R language. *Bioinformatics* **20**, 289–290 (2004).
37. Paradis, E. & Schliep, K. ape 5.0: An environment for modern phylogenetics and evolutionary analyses in R. *Bioinformatics* **35**, 526–528 (2018).
38. Yu, G., Smith, D. K., Zhu, H., Guan, Y. & Lam, T.T.-Y. GGTREE: An R package for visualization and annotation of phylogenetic trees with their covariates and other associated data. *Methods Ecol. Evol.* **8**, 28–36 (2017).
39. Wickham, H. *et al. ggplot2: Create elegant data visualisations using the grammar of graphics* (2022). https://CRAN.R-project.org/package=ggplot2. R package version 3.4.0.
40. Pedersen, T. L. *ggforce: Accelerating 'ggplot2'* (2022). https://CRAN.R-project.org/package=ggforce. R package version 0.4.1.

41. Gearty, W., Jones, L. A. & Chamberlain, S. *rphylopic: An R package for accessing and plotting PhyloPic silhouettes* (2023). https://github.com/palaeoverse-community/rphylopic. R package version 1.0.0.
42. Garnier, S. *et al.* viridisLite: *Colorblind-friendly color maps (lite version)* (2022). https://CRAN.R-project.org/package=viridisLite. R package version 0.4.1.
43. Goloboff, P. A., De Laet, J., Ríos-Tamayo, D. & Szumik, C. A. A reconsideration of inapplicable characters, and an approximation with step-matrix recoding. *Cladistics* **37**, 596–629 (2021).
44. Hopkins, M. J. & St. John, K. Incorporating hierarchical characters into phylogenetic analysis. *Syst. Biol.* **70**, 1163–1180 (2021).
45. Tarasov, S. New phylogenetic Markov models for inapplicable morphological characters. *Syst. Biol.*, syad005. https://doi.org/10.1093/sysbio/syad005 (2023).
46. Betancur-R, R. *et al.* Phylogenomic incongruence, hypothesis testing, and taxonomic sampling: The monophyly of characiform fishes. *Evolution* **73**, 329–345 (2019).
47. Felsenstein, J. Cases in which parsimony or compatibility methods will be positively misleading. *Syst. Zool.* **27**, 401–410 (1978).
48. Wright, A. M. & Hillis, D. M. Bayesian analysis using a simple likelihood model outperforms parsimony for estimation of phylogeny from discrete morphological data. *PLoS ONE* **9**, e109210 (2014).
49. O'Reilly, J. E. *et al.* Bayesian methods outperform parsimony but at the expense of precision in the estimation of phylogeny from discrete morphological data. *Biol. Lett.* **12**, 20160081 (2016).
50. O'Reilly, J. E., Puttick, M. N., Pisani, D. & Donoghue, P. C. J. Probabilistic methods surpass parsimony when assessing clade support in phylogenetic analyses of discrete morphological data. *Palaeontology* **61**, 105–118 (2018).
51. Puttick, M. N., O'Reilly, J. E., Pisani, D. & Donoghue, P. C. J. Probabilistic methods outperform parsimony in the phylogenetic analysis of data simulated without a probabilistic model. *Palaeontology* **62**, 1–17 (2019).
52. Minh, B. Q. *et al.* IQ-TREE 2: New models and efficient methods for phylogenetic inference in the genomic era. *Mol. Biol. Evol.* **37**, 1530–1534 (2020).
53. Lewis, P. O. A likelihood approach to estimating phylogeny from discrete morphological character data. *Syst. Biol.* **50**, 913–925 (2001).
54. Braun, E. L. & Kimball, R. T. Polytomies, the power of phylogenetic inference, and the stochastic nature of molecular evolution: A comment on Walsh et al. (1999). *Evolution* **55**, 1261–1263 (2001).
55. Poe, S. & Chubb, A. L. Birds in a bush: Five genes indicate explosive evolution of avian orders. *Evolution* **58**, 404–415 (2004).
56. Verbruggen, H. *et al.* Data mining approach identifies research priorities and data requirements for resolving the red algal tree of life. *BMC Evol. Biol.* **10**, 16 (2010).
57. Sayyari, E. & Mirarab, S. Testing for polytomies in phylogenetic species trees using quartet frequencies. *Genes* **9**, 132 (2018).
58. Haag, J., Höhler, D., Bettisworth, B. & Stamatakis, A. From easy to hopeless–Predicting the difficulty of phylogenetic analyses. *Mol. Biol. Evol.* **39**, msac254 (2022).
59. Robinson, D. F. & Foulds, L. R. Comparison of phylogenetic trees. *Math. Biosci.* **53**, 131–147 (1981).
60. Morel, B. *et al.* Phylogenetic analysis of SARS-CoV-2 data is difficult. *Mol. Biol. Evol.* **38**, 1777–1791 (2020).
61. Kishino, H., Miyata, T. & Hasegawa, M. Maximum likelihood inference of protein phylogeny and the origin of chloroplasts. *J. Mol. Evol.* **31**, 151–160 (1990).
62. Kishino, H. & Hasegawa, M. Evaluation of the maximum likelihood estimate of the evolutionary tree topologies from DNA sequence data, and the branching order in Hominoidea. *J. Mol. Evol.* **29**, 170–179 (1989).
63. Shimodaira, H. & Hasegawa, M. Multiple comparisons of log-likelihoods with applications to phylogenetic inference. *Mol. Biol. Evol.* **16**, 1114–1114 (1999).
64. Shimodaira, H. An approximately unbiased test of phylogenetic tree selection. *Syst. Biol.* **51**, 492–508 (2002).
65. Strimmer, K. & Rambaut, A. Inferring confidence sets of possibly misspecified gene trees. *Proc. R. Soc. B Biol. Sci.* **269**, 137–142 (2002).
66. Revell, L. J. phytools: An R package for phylogenetic comparative biology (and other things). *Methods Ecol. Evol.* **3**, 217–223 (2012).
67. Smith, M. R. *TreeTools: Create, modify and analyse phylogenetic trees* (2019). https://CRAN.R-project.org/package=TreeTools. R package version 1.9.0.
68. Hoang, D. T., Chernomor, O., von Haeseler, A., Minh, B. Q. & Vinh, L. S. UFBoot2: Improving the ultrafast bootstrap approximation. *Mol. Biol. Evol.* **35**, 518–522 (2017).
69. Menzel, U. *EMT: Exact multinomial test: Goodness-of-fit test for discrete multivariate data* (2021). https://CRAN.R-project.org/package=EMT. R package version 1.2.
70. Francis, W. R. & Canfield, D. E. Very few sites can reshape the inferred phylogenetic tree. *PeerJ* **8**, e8865 (2020).
71. Schliep, K. P. phangorn: Phylogenetic analysis in R. *Bioinformatics* **27**, 592–593 (2010).
72. Langer, M. C., McPhee, B. W., de Almeida Marsola, J. C., Roberto-da Silva, L. & Cabreira, S. F. Anatomy of the dinosaur *Pampadromaeus barberenai* (Saurischia–Sauropodomorpha) from the Late Triassic Santa Maria Formation of southern Brazil. *PLoS ONE* **14**, e0212543 (2019).
73. Wickham, H. & Seidel, D. *scales: Scale functions for visualization* (2022). https://CRAN.R-project.org/package=scales. R package version 1.2.1.
74. Wilke, C. O. *cowplot: Streamlined plot theme and plot annotations for 'ggplot2'* (2020). https://CRAN.R-project.org/package=cowplot. R package version 1.1.1.
75. Pedersen, T. L. *patchwork: The composer of plots* (2022). https://CRAN.R-project.org/package=patchwork. R package version 1.1.2.
76. Campitelli, E. *ggnewscale: Multiple fill and colour scales in 'ggplot2'* (2022). https://CRAN.R-project.org/package=ggnewscale. R package version 0.4.8.
77. Wickham, H., François, R., Henry, L., Müller, K. & Vaughan, D. *dplyr: A grammar of data manipulation* (2022). https://CRAN.R-project.org/package=dplyr. R package version 1.0.10.
78. Butler, R. J., Upchurch, P. & Norman, D. B. The phylogeny of the ornithischian dinosaurs. *J. Syst. Palaeontol.* **6**, 1–40 (2008).
79. Giribet, G. A new dimension in combining data? The use of morphology and phylogenomic data in metazoan systematics. *Acta Zool.* **91**, 11–19 (2010).
80. Gee, B. M. Returning to the roots: Resolution, reproducibility, and robusticity in the phylogenetic inference of Dissorophidae (Amphibia: Temnospondyli). *PeerJ* **9**, e12423 (2021).
81. Eliason, C. M., Edwards, S. V. & Clarke, J. A. phenotools: An R package for visualizing and analysing phenomic datasets. *Methods Ecol. Evol.* **10**, 1393–1400 (2019).
82. Porto, D. S. *et al.* Assessing Bayesian phylogenetic information content of morphological data using knowledge from anatomy ontologies. *Syst. Biol.* **71**, 1290–1306 (2022).
83. Poe, S. & Wiens, J. J. Character selection and the methodology of morphological phylogenetics. In *Phylogenetic Analysis of Morphological Data* (ed. Wiens, J. J.) 20–36 (Smithsonian Institution Press, 2000).
84. Höhna, S. *et al.* RevBayes: Bayesian phylogenetic inference using graphical models and an interactive model-specification language. *Syst. Biol.* **65**, 726–736 (2016).
85. Bouckaert, R. *et al.* BEAST 2.5: An advanced software platform for Bayesian evolutionary analysis. *PLoS Comput. Biol.* **15**, e1006650 (2019).

86. Regalado Fernández, O. R. & Werneburg, I. A new massopodan sauropodomorph from Trossingen Formation (Germany) hidden as 'Plateosaurus' for 100 years in the historical Tübingen collection. *Vertebr. Zool.* **72**, 771–822 (2022).
87. Wu, R., Pisani, D. & Donoghue, P. C. J. The unbearable uncertainty of panarthropod relationships. *Biol. Lett.* **19**, 20220497 (2023).
88. Huelsenbeck, J. P., Larget, B., Miller, R. E. & Ronquist, F. Potential applications and pitfalls of Bayesian inference of phylogeny. *Syst. Biol.* **51**, 673–688 (2002).
89. Holder, M. & Lewis, P. O. Phylogeny estimation: Traditional and Bayesian approaches. *Nat. Rev. Genet.* **4**, 275–284 (2003).
90. Kass, R. E. & Raftery, A. E. Bayes factors. *J. Am. Stat. Assoc.* **90**, 773–795 (1995).
91. Suchard, M. A., Weiss, R. E. & Sinsheimer, J. S. Models for estimating Bayes factors with applications to phylogeny and tests of monophyly. *Biometrics* **61**, 665–673 (2005).
92. Bergsten, J., Nilsson, A. N. & Ronquist, F. Bayesian tests of topology hypotheses with an example from diving beetles. *Syst. Biol.* **62**, 660–673 (2013).
93. Fourment, M. *et al.* 19 dubious ways to compute the marginal likelihood of a phylogenetic tree topology. *Syst. Biol.* **69**, 209–220 (2019).
94. O'Connor, P. M. *et al.* Late CRETACEOUS bird from Madagascar reveals unique development of beaks. *Nature* **588**, 272–276 (2020).
95. Felice, R. N. *et al.* Decelerated dinosaur skull evolution with the origin of birds. *PLoS Biol.* **18**, e3000801 (2020).
96. Müller, R. T. & Garcia, M. S. A paraphyletic 'Silesauridae' as an alternative hypothesis for the initial radiation of ornithischian dinosaurs. *Biol. Lett.* **16**, 20200417 (2020).
97. Castiglione, S., Serio, C., Mondanaro, A., Melchionna, M. & Raia, P. Fast production of large, time-calibrated, informal supertrees with `tree.merger`. *Palaeontology* **65**, e12588 (2022).
98. Baron, M. G. The effect of character and outgroup choice on the phylogenetic position of the Jurassic dinosaur *Chilesaurus diegosaurezi*. *Palaeoworld* https://doi.org/10.1016/j.palwor.2022.12.001 (2022).
99. Ezcurra, M. D. A new early dinosaur (Saurischia: Sauropodomorpha) from the Late Triassic of Argentina: A reassessment of dinosaur origin and phylogeny. *J. Syst. Palaeontol.* **8**, 371–425 (2010).
100. Cabreira, S. F. *et al.* A unique Late Triassic dinosauromorph assemblage reveals dinosaur ancestral anatomy and diet. *Curr. Biol.* **26**, 3090–3095 (2016).
101. Norman, D. B., Baron, M. G., Garcia, M. S. & Müller, R. T. Taxonomic, palaeobiological and evolutionary implications of a phylogenetic hypothesis for Ornithischia (Archosauria: Dinosauria). *Zool. J. Linn. Soc.* **196**, 1273–1309 (2022).
102. Hendrickx, C. *et al.* Morphology and distribution of scales, dermal ossifications, and other non-feather integumentary structures in non-avialan theropod dinosaurs. *Biol. Rev.* **97**, 960–1004 (2022).

Acknowledgements

Author contributions

D.Č. and A.L.S. conceived the study, performed the phylogenetic analyses, and interpreted the results. A.L.S. compiled the data and carried out detailed examination of character coding. D.Č. designed the methodology, scripted tree post-processing and statistical analyses, and generated the figures. D.Č. wrote the paper with input from A.L.S.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-023-35784-3.

Correspondence and requests for materials should be addressed to D.Č.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.