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

David Černý^{1,*} and Ashley L. Simonoff^{1,†}

¹ *Department of the Geophysical Sciences, University of Chicago, 5734 South Ellis Avenue, Chicago, IL 60637, USA*

* *Corresponding author. Email: david.cerny1@email.com*

† *Email: ashleysimonoff@gmail.com*

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 both the original and rescored dataset, we examine the global support for alternative hypotheses as well as the distribution of phylogenetic signal among individual characters. 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 (1887) 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 (Romer, 1956; Charig et al., 1965; Reig, 1970; Thulborn, 1975; Charig, 1982), 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) (Bakker and Galton, 1974; Bonaparte, 1976; Cooper, 1981). In the mid-1980s, the latter scenario was formalized as a phylogenetic hypothesis linking Ornithischia and Sauropodomorpha to the exclusion of Theropoda (Paul, 1984, 1988; Sereno, 1984) in a group variously termed Ornithischiformes (Cooper, 1985) or Phytodinosauria (Bakker, 1986) (Fig. 1). However, this hypothesis was abandoned after the first rigorous application of algorithmic phylogenetics to nonavian dinosaurs by Gauthier (1986), 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 (Novas, 1996; Benton, 2004; Nesbitt, 2011).

Using a large new dataset, Baron et al. (2017a; 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 (1870) name Ornithoscelida (Fig. 1). Seven months later, Langer et al. (2017; 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* (Baron et al., 2017b; Baron, 2019) and *Chilesaurus* (Baron and Barrett, 2017; Müller et al., 2018; Baron and Barrett, 2018; Müller and Dias-da Silva, 2019), nor by the use of more sophisticated phylogenetic methods such as time-free Bayesian inference (Parry et al., 2017) and Bayesian tip-dating (Griffin et al., 2022).

55 Despite the considerable interest generated by BEA's contribution, there have been few attempts to
56 quantify the relative support for each of the candidate topologies, distinguish between low information
57 content and internal conflict, or identify the characters that may drive such conflict – lines of investigation
58 that are more typical of phylogenomics (Reddy et al., 2017; Shen et al., 2017; Pease et al., 2018) than
59 morphological phylogenetics. LEA took early steps in this direction by demonstrating that when their
60 data was analyzed in a parsimony framework, the reinstated Saurischia hypothesis was statistically
61 indistinguishable from either of the alternatives (Langer et al., 2017). However, subsequent studies have
62 mostly reverted to reporting a single point estimate of the phylogeny, without testing whether its support
63 significantly exceeded that of the next best hypothesis. The occasional attempts to use the number of
64 extra steps relative to the most parsimonious tree for this purpose (Baron et al., 2017a; Baron and
65 Barrett, 2017) suffer from the fact that this quantity has no statistical interpretation (Felsenstein, 2004).
66 Moreover, despite the acknowledged centrality of character scoring differences to the conflict among the
67 resulting topologies (Langer et al., 2017; Baron et al., 2017b; Müller et al., 2018; Baron and Barrett,
68 2018), only one study to date has attempted to determine which rescorings drove the difference between
69 the hypotheses of early dinosaur phylogeny favored by BEA's and LEA's datasets (Goloboff and Sereno,
70 2021), and its methodological scope was limited to parsimony.

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

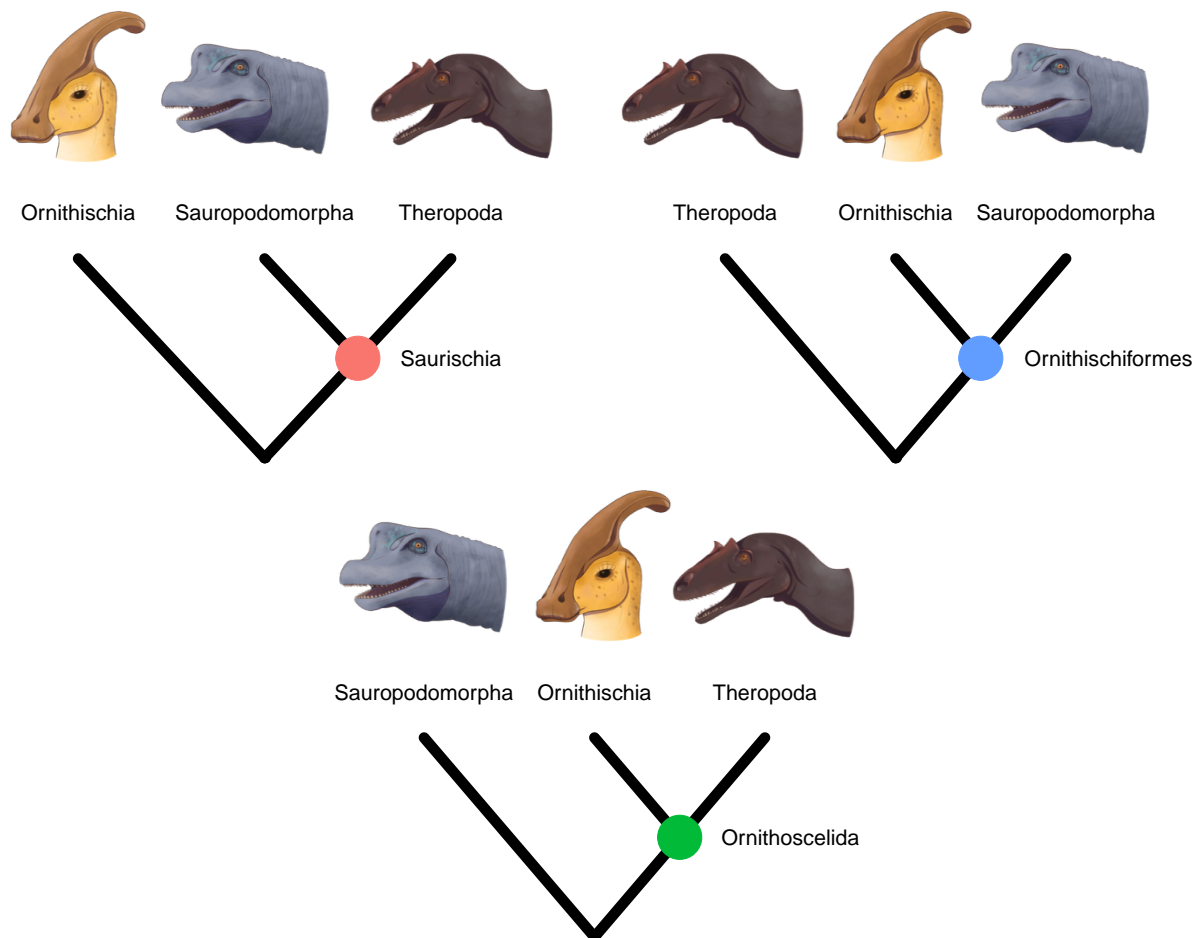


Fig. 1: Three alternative hypotheses of early dinosaur phylogeny evaluated in this study.

83 Results

84 *Maximum likelihood analyses*

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

	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.

97 Trees that were constrained to alternative early dinosaur topologies (Saurischia and Ornithischiformes
 98 for BEA, Saurischia and Ornithoscelida for LEA) exhibited log-likelihoods that fell well within the range
 99 yielded by the initial 100 unconstrained ML searches, suggesting that the three hypotheses were statisti-
 100 cally indistinguishable for both datasets. We formally corroborated this result using multiple likelihood-
 101 based topology tests, all of which indicated that none of the three topologies was significantly better or
 102 worse than the others (Tables 2, 3). The topological constraints were accommodated by the two datasets
 103 in markedly different ways. When their monophyly was enforced, Saurischia and Ornithischiformes were
 104 subtended by near-zero-length branches in the trees produced by the BEA dataset (Supplementary Figs.
 105 3, 4), indicating a lack of characters that could be convincingly interpreted as saurischian or ornithischi-
 106 form synapomorphies. For the LEA dataset, enforcing the monophyly of Saurischia yielded a topology
 107 that closely resembled the original parsimony estimate (cf. Supplementary Fig. 5 and Fig. 1 of Langer
 108 et al., 2017), while the tree optimized under an Ornithoscelida constraint showed an idiosyncratic topol-
 109 ogy that nested Ornithischia deep within theropods (Supplementary Fig. 6), reminiscent of hypotheses
 110 recently proposed by Baron (2019) to account for stratigraphic incongruence in ornithischian origins.

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 are indicated next to each test statistic.

111 *Character-wise support*

112 In both datasets, the distribution of character support for each of the three hypotheses was indistin-
 113 guishable from uniform, both across the matrix as a whole (multinomial test; BEA: $p = 0.063$, LEA:
 114 $p = 0.668$) and within most of the individual anatomical partitions (Fig. 2b,d). Despite the overall simi-
 115 larity of character support distributions between the two matrices, we found substantial differences at
 116 the level of individual characters. Only 70 out of the 447 overlapping non-constant characters ranked the
 117 three hypotheses in the same way in terms of their log-likelihoods, less than the one-sixth expected by
 118 chance. Similarly, only 145 characters preferred the same hypothesis, a number that is also indistin-
 119 guishable from the one-third expected by chance (multinomial test: $p = 0.726$). While the number of characters
 120 supporting Ornithoscelida decreased as a result of LEA's rescaling (from 174 to 147) and the number of
 121 characters supporting Saurischia marginally increased (from 141 to 143), even in the LEA dataset, more
 122 characters supported Ornithoscelida than Saurischia, although the difference was negligible (Fig. 2b,d).

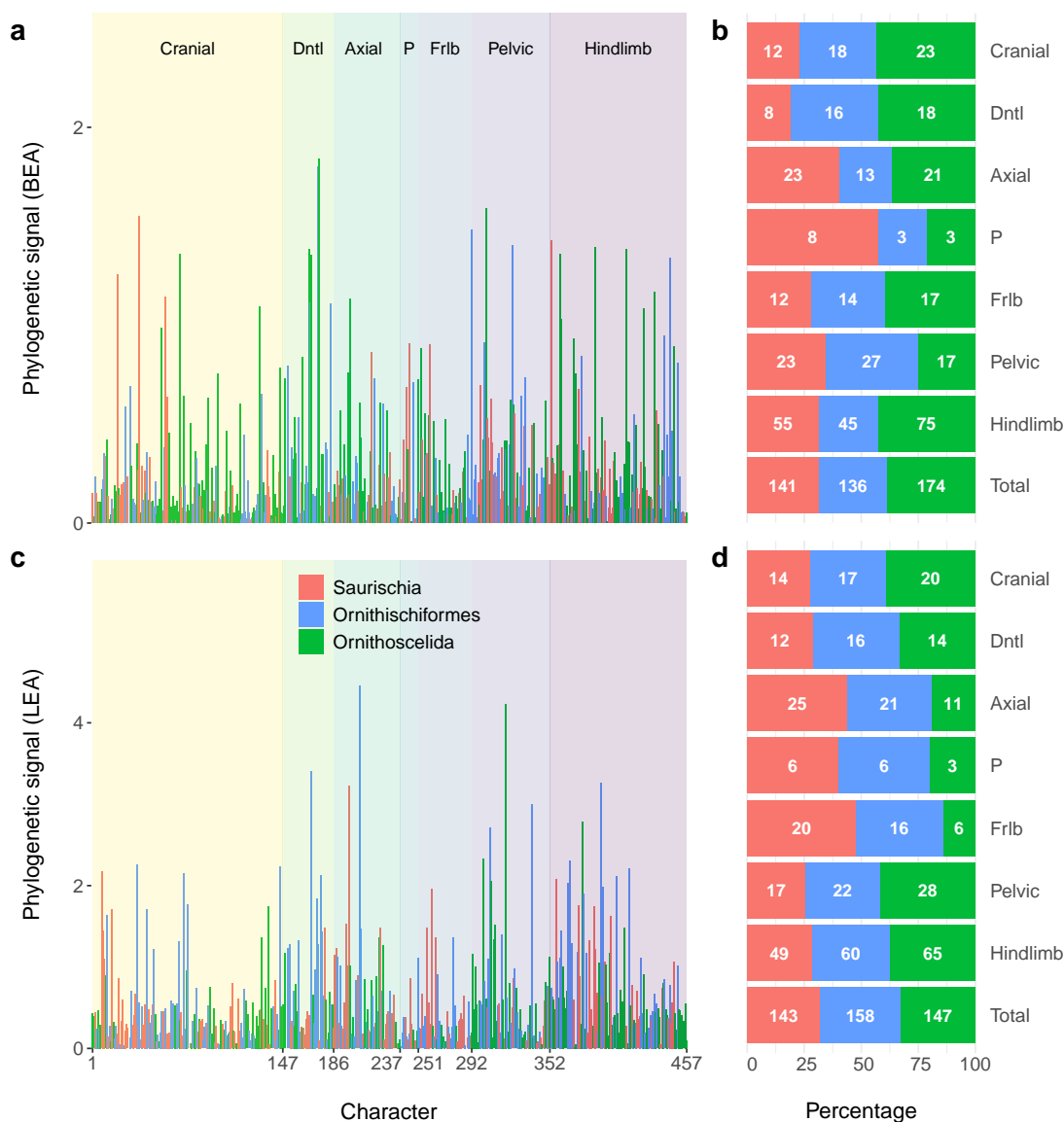


Fig. 2: Distribution of phylogenetic signal across the two datasets. Phylogenetic signal of every non-constant character from the BEA (a) and LEA (c) dataset is plotted against its placement in the matrix and colored by the preferred (highest-likelihood) hypothesis. The proportions and counts of characters preferring 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.

123 Examining support for competing hypotheses in terms of simple preference (i.e., by recording which
 124 tree yields the highest log-likelihood score for a given character) fails to account for the fact that most
 125 characters do not strongly prefer any of the three alternatives, rendering the resulting log-likelihood
 126 differences negligible and overly sensitive to minor branch length differences. Indeed, the character-wise
 127 log-likelihood difference (ΔCLS) between the best and second best hypotheses exceeds a threshold of 0.5

128 for less than 10% of characters in the BEA dataset and 30% of characters in the LEA dataset, although
 129 we note that these proportions are still substantially higher than is typical of the phylogenomic datasets
 130 for which the threshold was originally defined (Shen et al., 2017; Francis and Canfield, 2020). In the
 131 BEA dataset, the majority of these “strong” characters (29 out of 45) support Ornithoscelida, which is
 132 also the case for a plurality (63 out of 137) of such characters in the LEA dataset.

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.

133 To obtain a more fine-grained view of key characters driving the support for particular topologies, we
 134 employed the method suggested by Shen et al. (2017) and calculated the phylogenetic signal (PS) of each
 135 character as the mean of the absolute values of the three pairwise log-likelihood differences (Saurischia vs.
 136 Ornithischiformes, Saurischia vs. Ornithoscelida, Ornithischiformes vs. Ornithoscelida). Relative to the
 137 BEA dataset, the LEA dataset displays substantially higher mean (0.612 vs. 0.295) as well as maximum
 138 (4.454 vs. 1.840) phylogenetic signal values, consistent with its higher number of strong characters.
 139 The PS values of individual characters also differ considerably between the two datasets (Fig. 2a,c). In
 140 particular, among the characters with outlier PS values (more than three standard deviations above the
 141 mean), only one (character 169, serrations of maxillary and dentary teeth) is shared between the BEA
 142 matrix (in descending order of PS: 175, 174, 303, 37, 292, 353, 323, 387, 167, 411, 68, 360, 169, 444)
 143 and the LEA matrix (206, 318, 169, 391, 198, 338, 377, 306). Gradual removal of 1, 5, or 10 characters
 144 with the highest PS values or of all characters whose PS values represented outliers never caused either
 145 matrix to switch from the preferred topology to an alternative one (Supplementary Figs. 7–14). The
 146 exclusion of high-PS characters made little difference to the high statistical support for Ornithoscelida in
 147 the BEA dataset (UFBoot = 96–99%; Supplementary Figs. 7–10) but caused a drastic erosion of support
 148 for Ornithischiformes in the LEA dataset (Ornithischia + derived sauropodomorphs: UFBoot = 69–93%;
 149 Ornithischiformes proper + early sauropodomorphs: UFBoot = 52–79%; Supplementary Figs. 11–14).

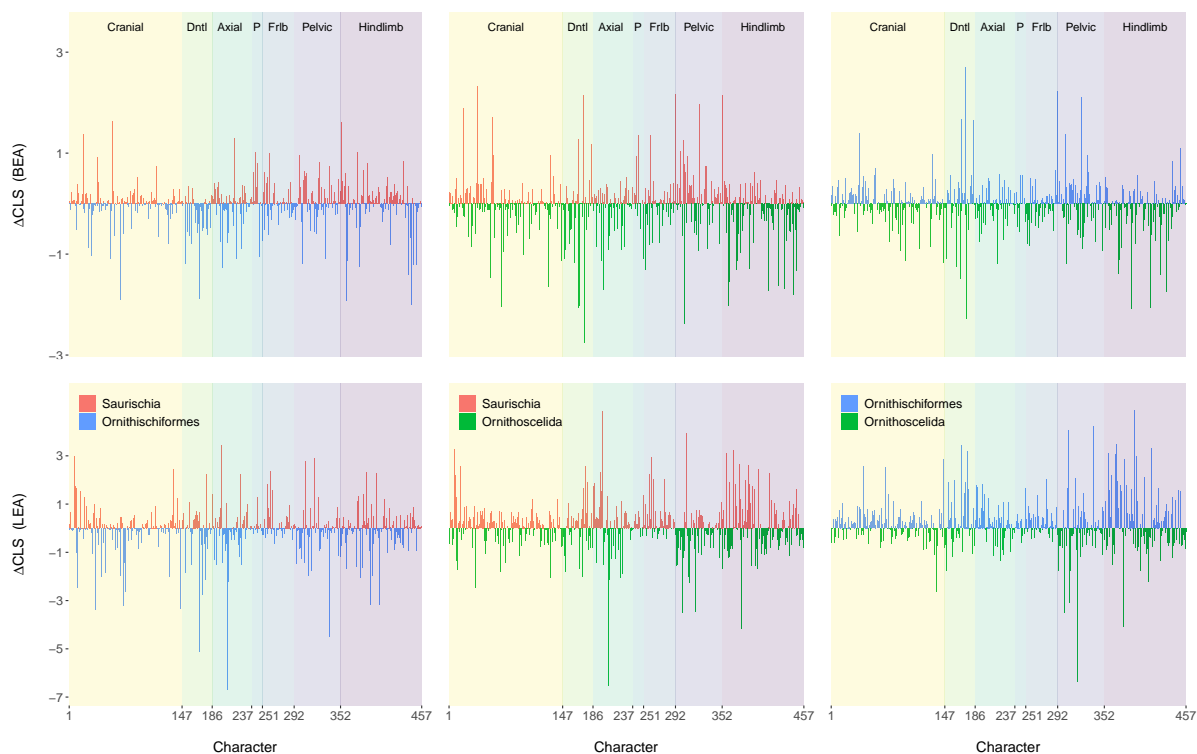


Fig. 3: Pairwise comparisons of character support for the three alternative early dinosaur topologies. For each pair of hypotheses, the difference in character-wise log-likelihood values (Δ CLS) of every non-constant character from the BEA (top) and LEA (bottom) datasets is plotted against its placement in the matrix. Anatomical region abbreviations as in Fig. 2.

150 To further distinguish among different ways in which a character can attain a high PS value, we
 151 separately compared each pair of hypotheses (Fig. 3), focusing on those characters that also emerged
 152 as outliers in at least two of the three pairwise comparisons. These corresponded to characters that
 153 strongly favored a particular hypothesis, strongly disfavored one, or both. The results reveal substantial
 154 conflict among the high-PS characters, as each matrix contains characters that both strongly favor and
 155 strongly disfavor its globally preferred topology (Ornithoscelida for BEA, Ornithischiformes for LEA)
 156 as well as one or both of its alternatives (Fig. 3, Table 4). Neither of the characters that strongly favor
 157 Ornithoscelida in the BEA dataset was among the 21 synapomorphies of this clade originally identified
 158 by Baron et al. (2017a), although one such synapomorphy (character 360, state 1: medial bowing of the
 159 femur forming a gentle curve) was found to strongly disfavor Saurischia in the present analysis (Table 4).
 160 Similarly, there is virtually no overlap between the characters strongly favoring or disfavoring one of
 161 the three hypotheses in the present study, and the “keystone” characters recently identified using a
 162 parsimony-based approach (Goloboff and Sereno, 2021).

163 *Rescoring of individual characters*

164 To determine which of LEA’s coding changes most contributed to the difference between the topologies
 165 yielded by the original and rescored datasets, we successively replaced the scoring of each character in
 166 either matrix by its scoring from the opposite matrix, and checked whether this change was sufficient
 167 for the re-estimated ML tree to switch to a different topology (Fig. 4). The procedure had little impact
 168 on the BEA dataset’s support for Ornithoscelida, which proved robust to the rescoring of any one of
 169 the 438 applicable characters and remained high on average (mean UFBoot = 98.8%; Fig. 4a). Only
 170 three characters (110, 114, 387) caused the support for Ornithoscelida to drop below the 95% threshold
 171 when changed to their scorings in the LEA dataset. Of these, none was optimized as an ornithoscelidan
 172 synapomorphy in BEA’s original analysis (Baron et al., 2017a), and while all favored Ornithoscelida over
 173 the alternatives in the original BEA matrix, only character 387 did so strongly ($\Delta\text{CLS} > 0.5$ relative to
 174 the next best hypothesis). In the LEA matrix, all three characters ranked Saurischia and Ornithoscelida
 175 as the best- and worst-supported hypothesis, respectively, but the log-likelihood difference between the
 176 two exceeded the 0.5 threshold only for character 387.

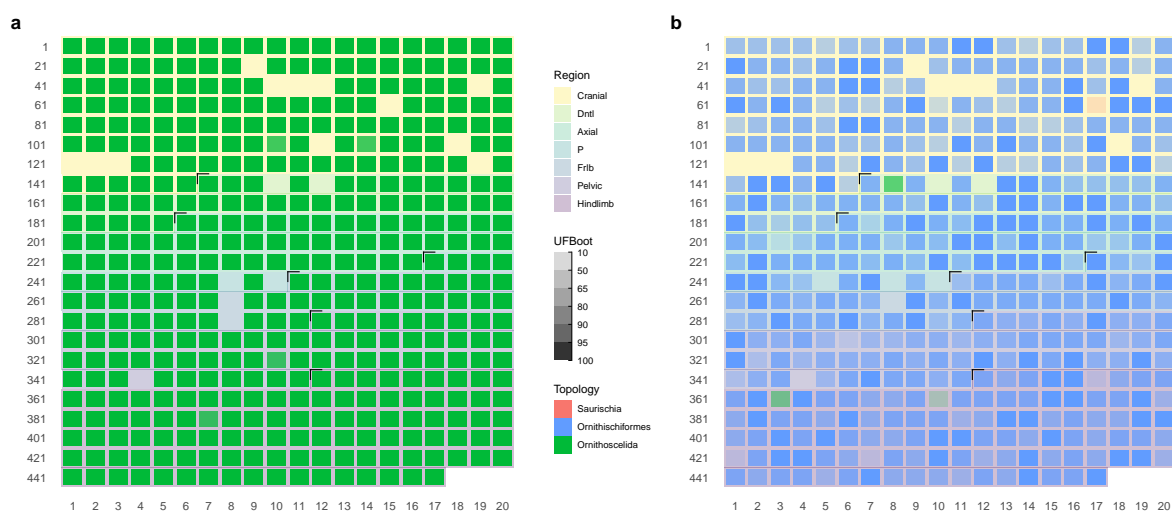


Fig. 4: Support for alternative topologies after rescoring each matrix one character at a time. Characters changed from BEA’s to LEA’s coding (a) or from LEA’s to BEA’s coding (b) are arranged by row following their placement in either dataset; black marks denote the boundaries between adjacent anatomical regions denoted by different background colors. 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.

177 In contrast, reverting four characters to their original scoring in the BEA matrix proved sufficient to
 178 flip the ML result from the LEA dataset (Ornithischiformes) either to Ornithoscelida (characters 148,
 179 363, 370) or to Saurischia (character 77) (Fig. 4b), producing highly idiosyncratic topologies (Supplemen-
 180 tary Figs. 15–18). In the ornithoscelidan trees, Ornithischia was deeply nested within theropods (with
 181 *Panguraptor* and *Zupaysaurus* consistently recovered closer to Ornithischia than to other theropods),
 182 similar to the results obtained from a constrained analysis of the unmodified LEA matrix (Supplemen-

tary 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 rescoring 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 (Baron et al., 2017a; see also Baron and Barrett, 2017). The only saurischian topology found was similarly unconventional and poorly supported (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 (Shen et al., 2017), 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, 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 hypothesis ($\Delta\text{CLS} > 0.5$) is far higher (10–30%) than typical for phylogenomic data (0.1–7%; Shen et al., 2017; Francis and Canfield, 2020). Similarly, although both matrices exhibit a highly uneven distribution of phylogenetic signal and contain several outlier character 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.

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 [$\Delta\text{CLS}(S, \text{Of})$, $\Delta\text{CLS}(S, \text{Os})$, $\Delta\text{CLS}(\text{Of}, \text{Os})$] 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.

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

231 First, we recommend re-examining the original dataset at a deeper level, as the recovery of divergent
232 yet statistically indistinguishable topologies may serve to highlight fundamental issues in the underlying
233 character data. Poorly formulated characters should ideally be redefined rather than simply rescored. For
234 example, LEA’s coding changes that caused character 174 (recurvature of maxillary and dentary teeth)
235 to lose the outlier status it originally had in the BEA dataset (Table 4) still took place in the framework
236 of the vague definition inherited from BEA, who in turn modified it from an even earlier study (Butler
237 et al., 2008). The character description provides no quantitative criterion for differentiating between teeth
238 possessing strong, weak, or no recurvature, allowing for more or less arbitrary coding changes. Indeed,
239 the scoring of *Efraasia* was changed by LEA from no recurvature to weak recurvature without explicit
240 justification or photographic evidence, and on the basis of the same published sources which BEA cited
241 in support of their own original coding. The problem of vague character descriptions and subjective
242 scoring decisions is widespread in the two matrices (e.g., characters 114, 216, 266, 337) and compounded
243 by a number of additional issues, some of which were noted by BEA and LEA themselves. These include
244 multiple instances of scoring taxa for characters that cannot be ascertained from their known material
245 (Baron et al., 2017b) and character non-independence (Langer et al., 2017). While the former issue
246 represents a fundamental problem with the data and a potentially serious biasing factor, especially when it
247 stems from coding taxa based on assumed rather than observed morphologies (Giribet, 2010; Gee, 2021),
248 the latter might be mitigated at the methodological level. Frameworks capable of dealing with hierarchical
249 and correlated characters are under development (Tarasov, 2022), as are ontology-based methods for their
250 semi-automated detection and characterization (Eliason et al., 2019; Porto et al., 2022). More advanced
251 methods may ultimately also alleviate other problems with existing matrices. The BEA and LEA datasets
252 contain instances of problematic state delimitation that make it impossible to assign certain morphologies
253 to any existing state (e.g., character 28, length of antorbital fenestra equal to 10–15% of skull length), or
254 merge distinct morphologies into a single state (e.g., character 77, state 0: paraquadratic foramen small
255 or absent). Both exemplify a more general problem, namely the arbitrary discretization of characters
256 that would be more naturally treated as continuous (Poe and Wiens, 2000). This ubiquitous feature of
257 morphological phylogenetic datasets reflects long-standing methodological limitations, which have only
258 recently been overcome by phylogenetic software packages that simultaneously implement models for
259 discrete and continuous characters, making it possible to combine both types of data in a single analysis
260 (Höhna et al., 2016; Bouckaert et al., 2019).

261 Second, instead of attempting to rescore or redefine an entire dataset in an indiscriminate “shotgun”
262 approach, it is prudent to determine which characters are responsible for the signal in that dataset’s
263 results. While LEA re-examined the putative ornithoscelidan synapomorphies identified by BEA at an
264 admirable level of detail (Langer et al., 2017), both previous findings (Goloboff and Sereno, 2021) and our
265 own results demonstrate that the characters supporting a particular topology do not always coincide with
266 the characters that map as the synapomorphies of that topology’s focal clade. Methods for identifying
267 such critical characters are now available in parsimony (Goloboff and Sereno, 2021), maximum-likelihood
268 (Shen et al., 2017; Francis and Canfield, 2020), and Bayesian (Porto et al., 2022) frameworks, and can
269 be profitably used to narrow the focus of potential rescaling efforts, which often involve the time- and
270 labor-intensive process of gathering data from first-hand observations of multiple museum specimens. The
271 benefits of comprehensively revising a pre-existing character matrix have to be weighed against the costs
272 inherent to such an effort, as well as the risk of introducing new errors into the data. In morphological
273 phylogenetics, extensive reuse and iterative expansion of pre-existing matrices gives rise to complicated
274 dataset genealogies (Gee, 2021; Regalado Fernández and Werneburg, 2022) in which coding error can
275 propagate and compound over time. On the other hand, our results suggest that the effort invested into

276 the comprehensive rescoring of a large pre-existing dataset can be difficult to justify: the three hypotheses
277 of early dinosaur phylogeny were statistically indistinguishable based on the original BEA dataset, and
278 remain such after LEA’s extensive revision of it.

279 Third, we urge paleontologists to quantify the uncertainty associated with their phylogenetic hy-
280 potheses using well-characterized tools with a clear statistical interpretation. When alternative ways of
281 resolving a given branch are of interest, as in the controversy surrounding early dinosaur phylogeny, we
282 encourage the community to move beyond the mere reporting of a phylogenetic point estimate toward
283 explicitly testing it against the next best alternative. Following recent practice (Wu et al., 2023), we
284 used a variety of likelihood ratio tests (LRTs) to this end (Shimodaira and Hasegawa, 1999; Shimodaira,
285 2002; Strimmer and Rambaut, 2002), but other approaches are possible. Bayesian inference differs from
286 maximum likelihood in its treatment of nuisance parameters such as branch lengths or the parameters of
287 the substitution and rate heterogeneity models, which are jointly optimized with the parameter of inter-
288 est (topology) in maximum likelihood but marginalized over in Bayesian methods (Huelsenbeck et al.,
289 2002). Both approaches have their advantages (Holder and Lewis, 2003), and as a result, Bayes factors
290 – a Bayesian equivalent of LRTs, relying on marginal rather than joint likelihoods (Kass and Raftery,
291 1995) – can represent a useful alternative way of evaluating competing topologies (Suchard et al., 2005;
292 Bergsten et al., 2013). Although most of the best-performing marginal likelihood estimators are much
293 more computationally demanding than joint likelihood inference (Fourment et al., 2019), the minute size
294 of phylogenetic datasets employed by paleontologists makes their application relatively easy, and Bayes
295 factor topological comparisons are accordingly starting to see use in dinosaur phylogenetics (O’Connor
296 et al., 2020).

297 Taken together, our results suggest that large-scale dinosaur phylogeny is much more poorly under-
298 stood than commonly acknowledged. In particular, despite the widespread perception that saurischian
299 monophyly is challenged primarily by the recently proposed Ornithoscelida hypothesis (Felice et al.,
300 2020; Müller and Garcia, 2020; Castiglione et al., 2022), the earlier Ornithischiformes hypothesis re-
301 ceives comparable support, and is in fact weakly preferred when the LEA dataset is analyzed using
302 maximum likelihood (Table 3). Additional support for Ornithischiformes was also recently detected in
303 an independent dataset (Baron, 2022). Moreover, not only the three major hypotheses – Saurischia,
304 Ornithischiformes, and Ornithoscelida – but also a number of their variations nesting the ornithischians
305 deep within Sauropodomorpha or Theropoda (Supplementary Figs. 2, 6, 16–18) cannot be ruled out at
306 present. Indeed, the specific variation on the Ornithischiformes topology recovered in this study shows
307 the ornithischians to be not just sister to, but rather nested within Sauropodomorpha, with a clade
308 of Carnian to early Norian sauropodomorphs (approximately corresponding to the Guaibasauridae of
309 Ezcurra, 2010 or the Saturnaliidae of Langer et al., 2019) branching off before the ornithischians. This
310 result is consistent with a phylogeny inferred from the LEA dataset by Parry et al. (2017) using time-
311 free Bayesian inference, showing that the two model-based methods yield topologies that are much more
312 similar to each other than to those favored by parsimony. By positing an early divergence of the early
313 Late Triassic sauropodomorphs, this scenario helps reduce the temporal gap between the first appear-
314 ance of Ornithischia and its sister group (Baron, 2019), and shows remarkable congruence with the early
315 suggestions that the ornithischians may have arisen from within “prosauropods” (Bakker and Galton,
316 1974; Bonaparte, 1976; Cooper, 1981; Bakker, 1986).

317 Our findings suggest that higher-level dinosaur interrelationships represent a phylogenetic problem
318 of considerable difficulty that is unlikely to be conclusively resolved by minor additions and superficial
319 modifications to the datasets currently in use. While our investigation was limited to two such datasets
320 (Baron et al., 2017a; Langer et al., 2017), there are few reasons to believe that other matrices currently
321 employed by dinosaur paleontologists are free of the problems identified here. Indeed, the main alternative
322 to the datasets examined here has repeatedly lent support to yet another nonstandard hypothesis nesting
323 the putatively non-dinosaurian Silesauridae within Ornithischia (Cabreira et al., 2016; Müller and Garcia,
324 2020; Norman et al., 2022), indicating that the number of plausible early dinosaur phylogenies proliferates
325 even further when not only the three major clades, but also species-poor lineages such as Herrerasauridae
326 and Silesauridae are taken into consideration. As a result, the increasingly common practice of repeating
327 comparative analyses under the Saurischia and Ornithoscelida topologies (Felice et al., 2020; Castiglione
328 et al., 2022; Hendrickx et al., 2022) most likely severely understates the uncertainty associated with early
329 dinosaur phylogeny, potentially leading to biased or overconfident conclusions.

330 **Methods**

331 *Data*

332 Our analyses were performed on the original character matrices used by BEA and LEA, whose properties
333 are summarized in Table 1. Both were obtained from Graeme T. Lloyd’s database of previously pub-

lished character matrices with standardized formatting (<http://graemetlloyd.com/matrdino.html>; last accessed May 1, 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 the 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; Goloboff and Sereno, 2021) 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 Goloboff et al., 2021; Hopkins and St. John, 2021; Tarasov, 2022). 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 (Betancur-R. et al., 2019). 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 (Felsenstein, 1978), including in the context of morphological phylogenetics (Wright and Hillis, 2014; O’Reilly et al., 2016, 2018; Puttick et al., 2019). 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 (Minh et al., 2020) 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 Mk model (Lewis, 2001) 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 (Lewis, 2001) 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 (Braun and Kimball, 2001; Poe and Chubb, 2004; Verbruggen et al., 2010); however, these often rely on models that are inapplicable to morphological evolution, such as the multispecies coalescent (Sayyari and Mirarab, 2018). 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. (2022):

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

where \bar{d}_{all} is the average pairwise normalized Robinson-Foulds (RF; Robinson and Foulds, 1981) distance between the $n_{\text{all}} = 100$ inferred trees, n'_{all} is the number of unique topologies among the 100 ML trees, \bar{d}_{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. (2020), 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

388 implemented in IQ-TREE. These were conducted using 10,000 approximate-bootstrap replicates (`-zb`
389 `10000 -zw -au`) generated by the resampling estimated log-likelihood method (Kishino et al., 1990) and
390 included the Kishino-Hasegawa test (Kishino and Hasegawa, 1989), the unweighted and weighted SH test
391 (Shimodaira and Hasegawa, 1999), the approximately unbiased test (Shimodaira, 2002), and expected
392 likelihood weights (Strimmer and Rambaut, 2002). Postprocessing was carried out in the R statistical
393 computing environment (R Core Team, 2021) using the packages *phytools* (Revell, 2012), *TreeTools*
394 (Smith, 2019), and their respective dependencies.

395 After the difficulty assessment, we performed one more round of unconstrained ML searches consisting
396 of 10 runs (each with 100 starting trees). The overall ML estimate was obtained by selecting the best-
397 scoring tree from the pooled sample of the 100 exploratory and 10 final runs. To assess clade support,
398 we additionally performed ultrafast bootstrap approximation (UFBoot) with 1000 replicates (`-B 1000`),
399 either after the fact (if the best-scoring tree was found during the exploratory runs) or simultaneously with
400 the main tree search (for the last 10 runs). In addition to being relatively robust to model misspecification,
401 the ultrafast bootstrap is less biased than the standard nonparametric bootstrap (Hoang et al., 2017),
402 and we also found it to be more numerically stable. Finally, we carried out topologically constrained
403 analyses enforcing those hypotheses that were not supported by the unconstrained tree. Similar to the
404 main analyses, each consisted of 10 runs performed simultaneously with 1000 UFBoot replicates.

405 *a. Character-wise support*

406 We followed the protocol of Shen et al. (2017) to assess how support for the three competing hypotheses
407 ($S = \text{Saurischia}$, $Of = \text{Ornithischiformes}$, $Os = \text{Ornithoscelida}$) was distributed across characters in both
408 matrices, and to identify potential outliers. Using IQ-TREE, we estimated character-wise log-likelihood
409 values for the ML trees yielded by both unconstrained and topologically constrained searches (`-wslr`). We
410 first calculated the number of characters in either matrix that preferred a given hypothesis (i.e., yielded
411 the least negative log-likelihood value under it), and repeated this calculation for individual anatomical
412 partitions. Multinomial tests were carried out using the R package *EMT* (Menzel, 2021) to determine
413 whether the resulting distributions were significantly different from uniform. We further determined the
414 number of characters that ranked the three hypotheses in the same way in terms of their log likelihoods
415 between the two matrices. Next, we calculated the phylogenetic signal (PS) of the i -th character (C_i) in
416 a given matrix following Eq. (5) of Shen et al. (2017):

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

417 Since the observed distributions of PS values were heavier-tailed than those shown in Shen et al.
418 (2017), we used a different criterion to identify outliers, defining them as those characters whose PS
419 was more than three standard deviations above the mean (Francis and Canfield, 2020). To assess the
420 influence exerted by characters with strong phylogenetic signal on the resulting early dinosaur topologies,
421 we generated 8 subsampled datasets by removing 1, 5, and 10 characters with the highest PS values from
422 either matrix, as well as those characters whose PS values represented outliers according to the above
423 criterion (BEA: 14 characters, LEA: 8 characters). These subsampled matrices were then subject to
424 unconstrained ML searches under the same settings as the original datasets (10 runs of 100 starting trees
425 each + 1000 UFBoot replicates).

426 The definition of phylogenetic signal outlined above has recently been criticized for conflating cases
427 in which one topology is strongly favored relative to either of the alternatives, and cases in which
428 one topology is strongly disfavored relative to both alternatives that may nevertheless remain nearly
429 indistinguishable from each other (Francis and Canfield, 2020). To tease apart these two scenarios, we
430 further evaluated the difference in log-likelihood scores (ΔCLS) separately for each pair of hypotheses.
431 For example, following Eq. (2) of Shen et al. (2017), the support lent to Saurischia over Ornithoscelida
432 by the i -th character was calculated as:

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

433 We then applied the same criterion for outliers to absolute ΔCLS values derived from each pairwise
434 comparison, and identified those characters that represented outliers in at least two of the three com-
435 parisons. These corresponded to characters that either strongly favored or strongly disfavored a given
436 topology relative to both of the alternatives. We also applied the criterion suggested by Francis and
437 Canfield (2020) and identified those characters for which the log-likelihood difference between the best
438 and second best hypotheses exceeded 0.5.

439 *Rescoring of individual characters*

440 To identify characters whose rescoring may have had disproportionate impact on the resulting topology,
441 we ran further ML analyses on modified versions of both datasets in which we successively recoded
442 one character at a time to its scoring in the opposite dataset. We excluded from consideration those
443 characters whose coding did not change between the two matrices (see “Data”) as well as those that
444 would be rendered constant by reverting their scoring to that of the opposite dataset (see “Maximum
445 likelihood analyses”), resulting in a total of 879 analyses. All of these were conducted under the same
446 settings as the analyses of the original datasets (10 runs of 100 starting trees each + 1000 UFBoot
447 replicates). Using a custom R script employing the package *phangorn* (Schliep, 2010), we scored each
448 resulting ML topology for the recovery of Saurischia, Ornithischiformes, or Ornithoscelida, and extracted
449 the UFBoot value of whichever of these three clades was present in the tree. To facilitate this process, we
450 treated the names Saurischia, Ornithischiformes, and Ornithoscelida as referring to node-based clades,
451 operationally defining them as (*Dilophosaurus* + *Plateosaurus*), (*Scelidosaurus* + *Plateosaurus*), and
452 (*Dilophosaurus* + *Scelidosaurus*), respectively.

453 **Acknowledgments**

454 We thank Graham J. Slater for valuable comments and advice, and to Catherine Wray for the illustrations
455 used in Fig. 1. D.Č. further thanks Joseph F. Walker for the suggestion to apply phylogenomic methods
456 of investigating site-wise support for alternative topologies to morphological datasets.

457 **Competing interests**

458 The authors declare no competing interests.

459 **References**

- 460 Bakker, R. T. (1986). *The Dinosaur Heresies: New Theories Unlocking the Mystery of the Dinosaurs*
461 *and Their Extinction*. William Morrow & Co., New York.
- 462 Bakker, R. T. and Galton, P. M. (1974). Dinosaur monophyly and a new class of vertebrates. *Nature*,
463 248(5444):168–172.
- 464 Baron, M. G. (2019). *Pisanosaurus mertii* and the Triassic ornithischian crisis: could phylogeny offer a
465 solution? *Historical Biology*, 31(8):967–981.
- 466 Baron, M. G. (2022). The effect of character and outgroup choice on the phylogenetic position of the
467 Jurassic dinosaur *Chilesaurus diegosaurezi*. *Palaeoworld*. doi:10.1016/j.palwor.2022.12.001.
- 468 Baron, M. G. and Barrett, P. M. (2017). A dinosaur missing-link? *Chilesaurus* and the early evolution
469 of ornithischian dinosaurs. *Biology Letters*, 13(8):20170220.
- 470 Baron, M. G. and Barrett, P. M. (2018). Support for the placement of *Chilesaurus* within Ornithischia:
471 a reply to Müller *et al.* *Biology Letters*, 14(3):20180002.
- 472 Baron, M. G., Norman, D. B., and Barrett, P. M. (2017a). A new hypothesis of dinosaur relationships
473 and early dinosaur evolution. *Nature*, 543(7646):501–506.
- 474 Baron, M. G., Norman, D. B., and Barrett, P. M. (2017b). Baron *et al.* reply. *Nature*, 551(7678):E4–E5.
- 475 Benton, M. J. (2004). Origin and relationships of Dinosauria. In Weishampel, D. B., Dodson, P.,
476 and Osmólska, H., editors, *The Dinosauria, 2nd edition*, pages 7–19. University of California Press,
477 Berkeley, CA.
- 478 Bergsten, J., Nilsson, A. N., and Ronquist, F. (2013). Bayesian tests of topology hypotheses with an
479 example from diving beetles. *Systematic Biology*, 62(5):660–673.
- 480 Betancur-R., R., Arcila, D., Vari, R. P., Hughes, L. C., Oliveira, C., Sabaj, M. H., and Ortí, G. (2019).
481 Phylogenomic incongruence, hypothesis testing, and taxonomic sampling: The monophyly of characi-
482 form fishes. *Evolution*, 73(2):329–345.
- 483 Bonaparte, J. F. (1976). *Pisanosaurus mertii* Casamiquela and the origin of the Ornithischia. *Journal*
484 *of Paleontology*, 50(5):808–820.

- 485 Bouckaert, R., Vaughan, T. G., Barido-Sottani, J., Duchêne, S., Fourment, M., Gavryushkina, A., Heled,
486 J., Jones, G., Kühnert, D., De Maio, N., Matschiner, M., Mendes, F. K., Müller, N. F., Ogilvie, H. A.,
487 du Plessis, L., Poppinga, A., Rambaut, A., Rasmussen, D., Siveroni, I., Suchard, M. A., Wu, C.-H., Xie,
488 D., Zhang, C., Stadler, T., and Drummond, A. J. (2019). BEAST 2.5: An advanced software platform
489 for Bayesian evolutionary analysis. *PLOS Computational Biology*, 15(4):e1006650.
- 490 Braun, E. L. and Kimball, R. T. (2001). Polytomies, the power of phylogenetic inference, and the
491 stochastic nature of molecular evolution: A comment on Walsh et al. (1999). *Evolution*, 55(6):1261–
492 1263.
- 493 Butler, R. J., Upchurch, P., and Norman, D. B. (2008). The phylogeny of the ornithischian dinosaurs.
494 *Journal of Systematic Palaeontology*, 6(1):1–40.
- 495 Cabreira, S. F., Kellner, A. W. A., Dias-da Silva, S., Roberto da Silva, L., Bronzati, M., Marsola, J. C.
496 d. A., Müller, R. T., Bittencourt, J. d. S., Batista, B. J., Raugust, T., Carrilho, R., Brodt, A., and
497 Langer, M. C. (2016). A unique Late Triassic dinosauromorph assemblage reveals dinosaur ancestral
498 anatomy and diet. *Current Biology*, 26(22):3090–3095.
- 499 Castiglione, S., Serio, C., Mondanaro, A., Melchionna, M., and Raia, P. (2022). Fast production of large,
500 time-calibrated, informal supertrees with `tree.merger`. *Palaeontology*, 65(1):e12588.
- 501 Charig, A. J. (1982). Problems in dinosaur phylogeny: a reasoned approach to their attempted resolution.
502 *Geobios*, 15:113–126.
- 503 Charig, A. J., Attridge, J., and Crompton, A. W. (1965). On the origin of the sauropods and the
504 classification of the Saurischia. *Proceedings of the Linnean Society of London*, 176(2):197–221.
- 505 Cooper, M. R. (1981). The prosauropod dinosaur *Massospondylus carinatus* Owen from Zimbabwe: its
506 biology, mode of life and phylogenetic significance. *Occasional Papers of the National Museums and*
507 *Monuments of Rhodesia, series B, Natural sciences*, 6(10):689–840.
- 508 Cooper, M. R. (1985). A revision of the ornithischian dinosaur *Kangnasaurus coetzeei* Houghton, with
509 a classification of the Ornithischia. *Annals of the South African Museum*, 95(8):281–317.
- 510 Eliason, C. M., Edwards, S. V., and Clarke, J. A. (2019). `phenotools`: An R package for visualizing and
511 analysing phenomic datasets. *Methods in Ecology and Evolution*, 10(9):1393–1400.
- 512 Ezcurra, M. D. (2010). A new early dinosaur (Saurischia: Sauropodomorpha) from the Late Triassic
513 of Argentina: a reassessment of dinosaur origin and phylogeny. *Journal of Systematic Palaeontology*,
514 8(3):371–425.
- 515 Felice, R. N., Watanabe, A., Cuff, A. R., Hanson, M., Bhullar, B.-A. S., Rayfield, E. R., Witmer, L. M.,
516 Norell, M. A., and Goswami, A. (2020). Decelerated dinosaur skull evolution with the origin of birds.
517 *PLOS Biology*, 18(8):e3000801.
- 518 Felsenstein, J. (1978). Cases in which parsimony or compatibility methods will be positively misleading.
519 *Systematic Zoology*, 27(4):401–410.
- 520 Felsenstein, J. (2004). *Inferring Phylogenies*. Sinauer Associates, Sunderland, MA.
- 521 Fourment, M., Magee, A. F., Whidden, C., Bilge, A., Matsen, Frederick A, I., and Minin, V. N. (2019). 19
522 dubious ways to compute the marginal likelihood of a phylogenetic tree topology. *Systematic Biology*,
523 69(2):209–220.
- 524 Francis, W. R. and Canfield, D. E. (2020). Very few sites can reshape the inferred phylogenetic tree.
525 *PeerJ*, 8:e8865.
- 526 Gauthier, J. A. (1986). Saurischian monophyly and the origin of birds. *Memoirs of the California*
527 *Academy of Sciences*, 8:1–55.
- 528 Gee, B. M. (2021). Returning to the roots: resolution, reproducibility, and robusticity in the phylogenetic
529 inference of Dissorophidae (Amphibia: Temnospondyli). *PeerJ*, 9:e12423.
- 530 Giribet, G. (2010). A new dimension in combining data? The use of morphology and phylogenomic data
531 in metazoan systematics. *Acta Zoologica*, 91(1):11–19.
- 532 Goloboff, P. A., De Laet, J., Ríos-Tamayo, D., and Szumik, C. A. (2021). A reconsideration of inapplicable
533 characters, and an approximation with step-matrix recoding. *Cladistics*, 37(5):596–629.

- 534 Goloboff, P. A. and Sereno, P. C. (2021). Comparative cladistics: identifying the sources for differing phy-
535 logenetic results between competing morphology-based datasets. *Journal of Systematic Palaeontology*,
536 19(11):761–786.
- 537 Griffin, C. T., Wynd, B. M., Munyikwa, D., Broderick, T. J., Zondo, M., Tolan, S., Langer, M. C.,
538 Nesbitt, S. J., and Taruvunga, H. R. (2022). Africa’s oldest dinosaurs reveal early suppression of
539 dinosaur distribution. *Nature*, 609(7926):313–319.
- 540 Haag, J., Höhler, D., Bettisworth, B., and Stamatakis, A. (2022). From easy to hopeless—Predicting the
541 difficulty of phylogenetic analyses. *Molecular Biology and Evolution*, 39(12):msac254.
- 542 Hendrickx, C., Bell, P. R., Pittman, M., Milner, A. R. C., Cuesta, E., O’Connor, J., Loewen, M., Currie,
543 P. J., Mateus, O., Kaye, T. G., and Delcourt, R. (2022). Morphology and distribution of scales, der-
544 mal ossifications, and other non-feather integumentary structures in non-avian theropod dinosaurs.
545 *Biological Reviews*, 97(3):960–1004.
- 546 Hoang, D. T., Chernomor, O., von Haeseler, A., Minh, B. Q., and Vinh, L. S. (2017). UFBoot2: improving
547 the ultrafast bootstrap approximation. *Molecular Biology and Evolution*, 35(2):518–522.
- 548 Höhna, S., Landis, M. J., Heath, T. A., Boussau, B., Lartillot, N., Moore, B. R., Huelsenbeck, J. P.,
549 and Ronquist, F. (2016). RevBayes: Bayesian phylogenetic inference using graphical models and an
550 interactive model-specification language. *Systematic Biology*, 65(4):726–736.
- 551 Holder, M. and Lewis, P. O. (2003). Phylogeny estimation: traditional and Bayesian approaches. *Nature*
552 *Reviews Genetics*, 4(4):275–284.
- 553 Hopkins, M. J. and St. John, K. (2021). Incorporating hierarchical characters into phylogenetic analysis.
554 *Systematic Biology*, 70(6):1163–1180.
- 555 Huelsenbeck, J. P., Larget, B., Miller, R. E., and Ronquist, F. (2002). Potential applications and pitfalls
556 of Bayesian inference of phylogeny. *Systematic Biology*, 51(5):673–688.
- 557 Huxley, T. H. (1870). On the classification of the Dinosauria, with observations on the Dinosauria of the
558 Trias. *Quarterly Journal of the Geological Society*, 26(1-2):32–51.
- 559 Kass, R. E. and Raftery, A. E. (1995). Bayes factors. *Journal of the American Statistical Association*,
560 90(430):773–795.
- 561 Kishino, H. and Hasegawa, M. (1989). Evaluation of the maximum likelihood estimate of the evolutionary
562 tree topologies from DNA sequence data, and the branching order in Hominoidea. *J. Mol. Evol.*,
563 29(2):170–179.
- 564 Kishino, H., Miyata, T., and Hasegawa, M. (1990). Maximum likelihood inference of protein phylogeny
565 and the origin of chloroplasts. *Journal of Molecular Evolution*, 31(2):151–160.
- 566 Langer, M. C., Ezcurra, M. D., Rauhut, O. W. M., Benton, M. J., Knoll, F., McPhee, B. W., Novas, F. E.,
567 Pol, D., and Brusatte, S. L. (2017). Untangling the dinosaur family tree. *Nature*, 551(7678):E1–E3.
- 568 Langer, M. C., McPhee, B. W., Marsola, J. C. d. A., Roberto-da Silva, L., and Cabreira, S. F. (2019).
569 Anatomy of the dinosaur *Pampadromaeus barberenai* (Saurischia—Sauropodomorpha) from the Late
570 Triassic Santa Maria Formation of southern Brazil. *PLOS ONE*, 14(2):e0212543.
- 571 Lewis, P. O. (2001). A likelihood approach to estimating phylogeny from discrete morphological character
572 data. *Systematic Biology*, 50(6):913–925.
- 573 Menzel, U. (2021). *EMT: Exact multinomial test: goodness-of-fit test for discrete multivariate data*.
574 Comprehensive R Archive Network. R package version 1.2.
- 575 Minh, B. Q., Schmidt, H. A., Chernomor, O., Schrempf, D., Woodhams, M. D., von Haeseler, A., and
576 Lanfear, R. (2020). IQ-TREE 2: new models and efficient methods for phylogenetic inference in the
577 genomic era. *Molecular Biology and Evolution*, 37(5):1530–1534.
- 578 Morel, B., Barbera, P., Czech, L., Bettisworth, B., Hübner, L., Lutteropp, S., Serdari, D., Kostaki, E.-
579 G., Mamais, I., Kozlov, A. M., Pavlidis, P., Paraskevis, D., and Stamatakis, A. (2020). Phylogenetic
580 analysis of SARS-CoV-2 data is difficult. *Molecular Biology and Evolution*, 38(5):1777–1791.
- 581 Müller, R. T. and Dias-da Silva, S. (2019). Taxon sample and character coding deeply impact unstable
582 branches in phylogenetic trees of dinosaurs. *Historical Biology*, 31(8):1089–1092.

- 583 Müller, R. T. and Garcia, M. S. (2020). A paraphyletic ‘Silesauridae’ as an alternative hypothesis for
584 the initial radiation of ornithischian dinosaurs. *Biology Letters*, 16(8):20200417.
- 585 Müller, R. T., Pretto, F. A., Kerber, L., Silva-Neves, E., and Dias-da Silva, S. (2018). Comment on ‘A
586 dinosaur missing-link? *Chilesaurus* and the early evolution of ornithischian dinosaurs’. *Biology Letters*,
587 14(3):20170581.
- 588 Nesbitt, S. J. (2011). The early evolution of archosaurs: relationships and the origin of major clades.
589 *Bulletin of the American Museum of Natural History*, 352:1–292.
- 590 Norman, D. B., Baron, M. G., Garcia, M. S., and Müller, R. T. (2022). Taxonomic, palaeobiological
591 and evolutionary implications of a phylogenetic hypothesis for Ornithischia (Archosauria: Dinosauria).
592 *Zoological Journal of the Linnean Society*, 196(4):1273–1309.
- 593 Novas, F. E. (1996). Dinosaur monophyly. *Journal of Vertebrate Paleontology*, 16(4):723–741.
- 594 O’Connor, P. M., Turner, A. H., Groenke, J. R., Felice, R. N., Rogers, R. R., Krause, D. W., and
595 Rahantarisoa, L. J. (2020). Late cretaceous bird from Madagascar reveals unique development of
596 beaks. *Nature*, 588(7837):272–276.
- 597 O’Reilly, J. E., Puttick, M. N., Parry, L., Tanner, A. R., Tarver, J. E., Fleming, J., Pisani, D., and
598 Donoghue, P. C. J. (2016). Bayesian methods outperform parsimony but at the expense of precision
599 in the estimation of phylogeny from discrete morphological data. *Biology Letters*, 12(4):20160081.
- 600 O’Reilly, J. E., Puttick, M. N., Pisani, D., and Donoghue, P. C. J. (2018). Probabilistic methods sur-
601 pass parsimony when assessing clade support in phylogenetic analyses of discrete morphological data.
602 *Palaeontology*, 61(1):105–118.
- 603 Parry, L. A., Baron, M. G., and Vinther, J. (2017). Multiple optimality criteria support Ornithoscelida.
604 *Royal Society Open Science*, 4(10):170833.
- 605 Paul, G. S. (1984). The segnosaurian dinosaurs: relics of the prosauropod–ornithischian transition?
606 *Journal of Vertebrate Paleontology*, 4(4):507–515.
- 607 Paul, G. S. (1988). *Predatory Dinosaurs of the World: A Complete Illustrated Guide*. Simon and Schuster,
608 New York.
- 609 Pease, J. B., Brown, J. W., Walker, J. F., Hinchliff, C. E., and Smith, S. A. (2018). Quartet Sampling
610 distinguishes lack of support from conflicting support in the green plant tree of life. *American Journal*
611 *of Botany*, 105(3):385–403.
- 612 Poe, S. and Chubb, A. L. (2004). Birds in a bush: five genes indicate explosive evolution of avian orders.
613 *Evolution*, 58(2):404–415.
- 614 Poe, S. and Wiens, J. J. (2000). Character selection and the methodology of morphological phyloge-
615 netics. In Wiens, J. J., editor, *Phylogenetic Analysis of Morphological Data*, page 20–36. Smithsonian
616 Institution Press, Washington, DC.
- 617 Porto, D. S., Dahdul, W. M., Lapp, H., Balhoff, J. P., Vision, T. J., Mabee, P. M., and Uyeda, J.
618 (2022). Assessing Bayesian phylogenetic information content of morphological data using knowledge
619 from anatomy ontologies. *Systematic Biology*, 71(6):1290–1306.
- 620 Puttick, M. N., O’Reilly, J. E., Pisani, D., and Donoghue, P. C. J. (2019). Probabilistic methods outper-
621 form parsimony in the phylogenetic analysis of data simulated without a probabilistic model. *Palaeon-*
622 *tology*, 62(1):1–17.
- 623 R Core Team (2021). *R: A Language and Environment for Statistical Computing*. R Foundation for
624 Statistical Computing, Vienna, Austria.
- 625 Reddy, S., Kimball, R. T., Pandey, A., Hosner, P. A., Braun, M. J., Hackett, S. J., Han, K.-L., Harshman,
626 J., Huddleston, C. J., Kingston, S., Marks, B. D., Miglia, K. J., Moore, W. S., Sheldon, F. H., Witt,
627 C. C., Yuri, T., and Braun, E. L. (2017). Why do phylogenomic data sets yield conflicting trees? Data
628 type influences the avian tree of life more than taxon sampling. *Syst. Biol.*, 66(5):857–879.
- 629 Regalado Fernández, O. R. and Werneburg, I. (2022). A new massopodan sauropodomorph from Trossin-
630 gen Formation (Germany) hidden as ‘*Plateosaurus*’ for 100 years in the historical Tübingen collection.
631 *Vertebrate Zoology*, 72:771–822.

- 632 Reig, O. A. (1970). The Proterosuchia and the early evolution of the archosaurs; an essay about the
633 origin of a major taxon. *Bulletin of the Museum of Comparative Zoology*, 139(5):229–292.
- 634 Revell, L. J. (2012). phytools: an R package for phylogenetic comparative biology (and other things).
635 *Methods in Ecology and Evolution*, 3(2):217–223.
- 636 Robinson, D. F. and Foulds, L. R. (1981). Comparison of phylogenetic trees. *Mathematical Biosciences*,
637 53(1):131–147.
- 638 Romer, A. S. (1956). *Osteology of the Reptiles*. University of Chicago Press, Chicago, IL.
- 639 Sayyari, E. and Mirarab, S. (2018). Testing for polytomies in phylogenetic species trees using quartet
640 frequencies. *Genes*, 9(3):132.
- 641 Schliep, K. P. (2010). phangorn: phylogenetic analysis in R. *Bioinformatics*, 27(4):592–593.
- 642 Seeley, H. G. (1887). I. On the classification of the fossil animals commonly named Dinosauria. *Proceed-*
643 *ings of the Royal Society of London*, 43(258-265):165–171.
- 644 Sereno, P. C. (1984). The phylogeny of the Ornithischia: a reappraisal. In Reif, W.-E. and Westphal, F.,
645 editors, *Third Symposium on Mesozoic Terrestrial Ecosystems, Short Papers*, page 219–226. Attempto
646 Verlag, Tübingen.
- 647 Shen, X.-X., Hittinger, C. T., and Rokas, A. (2017). Contentious relationships in phylogenomic studies
648 can be driven by a handful of genes. *Nature Ecology & Evolution*, 1(5):0126.
- 649 Shimodaira, H. (2002). An approximately unbiased test of phylogenetic tree selection. *Syst. Biol.*,
650 51(3):492–508.
- 651 Shimodaira, H. and Hasegawa, M. (1999). Multiple comparisons of log-likelihoods with applications to
652 phylogenetic inference. *Mol. Biol. Evol.*, 16(8):1114–1114.
- 653 Smith, M. R. (2019). *TreeTools: create, modify and analyse phylogenetic trees*. Comprehensive R Archive
654 Network. R package version 1.9.0.
- 655 Strimmer, K. and Rambaut, A. (2002). Inferring confidence sets of possibly misspecified gene trees. *Proc.*
656 *R. Soc. Lond. B*, 269(1487):137–142.
- 657 Suchard, M. A., Weiss, R. E., and Sinsheimer, J. S. (2005). Models for estimating Bayes factors with
658 applications to phylogeny and tests of monophyly. *Biometrics*, 61(3):665–673.
- 659 Tarasov, S. (2022). New phylogenetic Markov models for inapplicable morphological characters. *bioRxiv*.
660 doi:10.1101/2021.04.26.441495.
- 661 Thulborn, R. A. (1975). Dinosaur polyphyly and the classification of archosaurs and birds. *Australian*
662 *Journal of Zoology*, 23(2):249–270.
- 663 Verbruggen, H., Maggs, C. A., Saunders, G. W., Le Gall, L., Yoon, H. S., and De Clerck, O. (2010).
664 Data mining approach identifies research priorities and data requirements for resolving the red algal
665 tree of life. *BMC Evolutionary Biology*, 10(1):16.
- 666 Wright, A. M. and Hillis, D. M. (2014). Bayesian analysis using a simple likelihood model outperforms
667 parsimony for estimation of phylogeny from discrete morphological data. *PLOS ONE*, 9(10):e109210.
- 668 Wu, R., Pisani, D., and Donoghue, P. C. J. (2023). The unbearable uncertainty of panarthropod rela-
669 tionships. *Biology Letters*, 19(1):20220497.