



Smith, T. J., Puttick, M. N., O'Reilly, J. E., Pisani, D., & Donoghue, P. C. J. (2021). Phylogenetic sampling affects evolutionary patterns of morphological disparity. *Palaeontology*, 64(6), 765-787.
<https://doi.org/10.1111/pala.12569>

Publisher's PDF, also known as Version of record

License (if available):
CC BY

Link to published version (if available):
[10.1111/pala.12569](https://doi.org/10.1111/pala.12569)

[Link to publication record in Explore Bristol Research](#)
PDF-document

This is the final published version of the article (version of record). It first appeared online via Wiley at <https://doi.org/10.1111/pala.12569> . Please refer to any applicable terms of use of the publisher.

University of Bristol - Explore Bristol Research

General rights

This document is made available in accordance with publisher policies. Please cite only the published version using the reference above. Full terms of use are available:
<http://www.bristol.ac.uk/red/research-policy/pure/user-guides/ebr-terms/>



PHYLOGENETIC SAMPLING AFFECTS EVOLUTIONARY PATTERNS OF MORPHOLOGICAL DISPARITY

by THOMAS J. SMITH¹ , MARK N. PUTTICK² , JOSEPH E. O'REILLY^{1,3} ,
DAVIDE PISANI¹  and PHILIP C. J. DONOGHUE¹ 

¹School of Earth Sciences, University of Bristol, Life Sciences Building, Tyndall Avenue, Bristol, BS8 1TQ, UK; tom.smith@bristol.ac.uk, joseph.oreilly@igmm.ed.ac.uk, davide.pisani@bristol.ac.uk, phil.donoghue@bristol.ac.uk

²Department of Biology & Biochemistry, Milner Centre for Evolution, University of Bath, Claverton Down, Bath, BA2 7AZ, UK; marknputtick@gmail.com

³Present address: MRC Human Genetics Unit, MRC, Institute of Genetics & Cancer, University of Edinburgh, Western General Hospital, Crewe Road, Edinburgh, EH4 2XU, UK

Typescript received 10 November 2020; accepted in revised form 26 June 2021

Abstract: Cladistic character matrices are routinely repurposed in analyses of morphological disparity. Unfortunately, the sampling of taxa and characters within such datasets reflects their intended application (to resolve phylogeny, rather than distinguish between phenotypes) resulting in tree shapes that often misrepresent broader taxonomic and morphological diversity. Here we use tree shape as a proxy to explore how sampling can affect perceptions of evolving morphological disparity. Through analyses of simulated and empirical data, we demonstrate that sampling can introduce biases in morphospace occupation between clades that are predicted by differences in tree symmetry and branch length distribution. Symmetrical trees with relatively long internal branches predict more expansive patterns of morphospace occupation. Conversely, asymmetrical trees with relatively short internal branches predict more compact distributions. Additionally, we find that long external branches predict greater phenotypic divergence by peripheral morphotypes.

Taken together, our results caution against the uncritical repurposing of cladistic datasets in disparity analyses. However, they also demonstrate that when morphological diversity is proportionately sampled, differences in tree shape between clades can speak to genuine differences in morphospace occupation. While cladistic datasets may serve as a useful starting point, disparity datasets must attempt to achieve uniformity in lineage sampling across time and topology. Only when all potential sources of bias are accounted for can evolutionary phenomena be distinguished from artefactual signals. It must be accepted that the non-uniformity of the fossil record may preclude representative sampling and, therefore, a faithful characterization of the evolution of morphological disparity.

Key words: disparity, morphology, simulation, sampling, topology, phylogeny.

ANALYSES of morphological disparity characterize the shape of life and provide a framework for uncovering the intrinsic and extrinsic causes of its evolution. Defined as a measure of the phenotypic variation exhibited within a sample of taxa, the versatility of morphological disparity methods has facilitated their application in a wide variety of contexts and to a diverse array of evolutionary questions (Foote 1997; Wills 2001; Erwin 2007; Guillerme *et al.* 2020a). Many contemporary disparity studies are based on repurposed discrete character cladistic datasets (see Lloyd 2016 for overview). This is potentially problematic as the intended outcomes of phylogenetic and disparity analyses differ; the former seeks to group taxa, while the latter aims to distinguish between them (Gerber 2019). Consequently, cladistic datasets often possess

characteristics that are beneficial in analyses of phylogeny but confounding in disparity analyses. These include biases in character sampling, which have been explored to some extent (Cisneros & Ruta 2010; Ruta & Wills 2016; Deline *et al.* 2018), but also taxon sampling (e.g. living vs fossil and ingroup vs outgroup taxa), which can result in biased tree shapes that do not faithfully represent diversity, either in terms of symmetry or branch lengths.

Here, we explore the impact of sampling on perceptions of morphological disparity and its evolution through the lens of tree shape. First, we summarize the history of simulations in analyses of morphological disparity. Then, categorical data are simulated along different trees to characterize the effects of symmetry and branch-length distribution, which we use independently

and in combination as proxies for different sampling practices. These simulated data are analysed, and the results compared to complementary analyses of empirical datasets. Both symmetry and branch length distribution predict the morphospace occupation of clades. Symmetrical trees with relatively long internal branches predict more expansive patterns of morphospace occupation, while asymmetrical trees with relatively short internal branches predict more compact distributions. Long external branches predict greater phenotypic divergence by peripheral morphotypes. Disparity-through-time analyses demonstrate that symmetry predicts the contributions of ancestral phenotypes to perceptions of evolving morphological disparity, whereas the branch length distribution of a tree has no such predictive value. However, the absence of a difference between data simulated along ultrametric and non-ultrametric trees in disparity-through-time analyses speaks to the potential utility of ancestral state estimation in analyses limited to extant taxa. Taken together, our results demonstrate that sampling practices adopted during the construction of cladistic datasets can bias the morphospaces they characterize. Therefore, these practices and the artefacts they introduce should be considered carefully before cladistic datasets are repurposed in analyses of morphological disparity. Researchers should seek to achieve faithful representations of morphological diversity through the augmentation or subsampling of existing cladistic matrices or, better still, the construction of datasets purpose-built for analyses of disparity.

SIMULATIONS IN ANALYSES OF DISPARITY

Simulated data offer several advantages over empirical datasets, making them ideally suited for the exploration of null trends in evolution (Barido-Sottani *et al.* 2020). First and foremost, the ability to simulate the evolutionary process iteratively in a static environment makes it possible to generate distributions that better represent the breadth of possible trajectories than single empirical datasets. Simulations can also be conducted while controlling for all potential sources of unwanted variation, thus promoting greater compatibility between each run than can be achieved with empirical data. Finally, simulated datasets can be generated with less time expenditure and manual effort than empirical data, allowing for the rapid generation of large quantities of data suitable for hypothesis testing.

Simulations have been employed previously in analyses of disparity, but they have been limited largely to continuous data, evaluating the contributions of phylogeny (Harmon *et al.* 2003), ecology (Green *et al.* 2011),

evolutionary rates (Foote 1991) and modes of extinction (Korn *et al.* 2013) to morphospace exploration. Simulation-based analyses of disparity have been used to test hypotheses of functional constraint (Niklas 1986), assess the impact of mass extinctions (Puttick *et al.*, 2020) and explore the origins of lineage clustering in morphospace (Pie & Weitz 2005). Simulated continuous data have also been used to assess the performance of dissimilarity indices in characterizing different aspects of morphospace (Ciampaglio *et al.* 2001; Guillerme *et al.* 2020*b*). For the purpose of such analyses, continuous data have been simulated under a variety of stochastic (Foote 1991; Ciampaglio *et al.* 2001; Harmon *et al.* 2003; Pie & Weitz 2005; Green *et al.* 2011; Korn *et al.* 2013; Guillerme *et al.* 2020*b*; Puttick *et al.* 2020) and analytical (Foote 1996*a*; Gerber *et al.* 2011) models. Probabilistic methods that permute an ancestral form in response to dynamic functional criteria have also been employed, although these have been limited to studies of plant evolution (Niklas 1986).

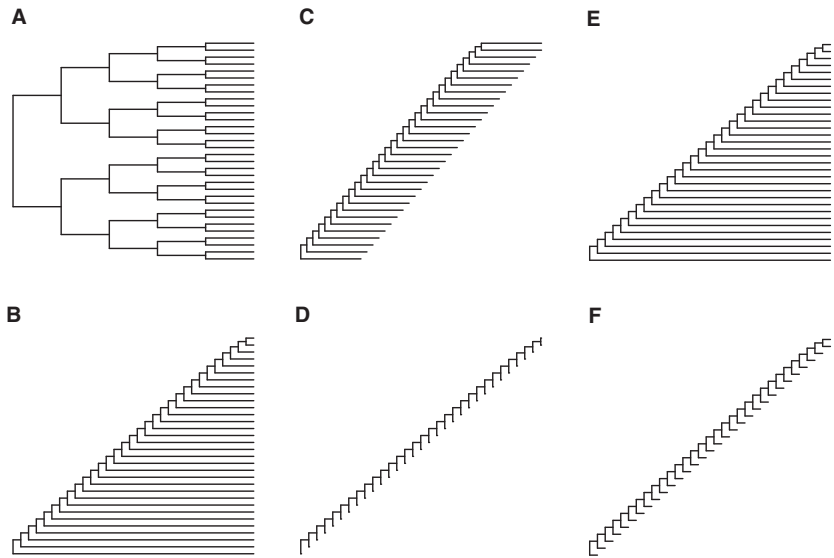
Simulation approaches have been adopted sparingly in analyses of disparity based on categorical data. Gavrilts (1999) employed a simple analytical model to explore the relationship between taxonomic and morphological diversification, while Ciampaglio *et al.* (2001) simulated discrete character data using a random walk to assess the relative performance of different indices of disparity. Sutherland *et al.* (2019) simulated random data to explore the relationship between missing data and morphospace occupation. Finally, Schaeffer *et al.* (2020) and Ferron *et al.* (2020) evolved characters along a phylogeny under Brownian motion to derive null expectations for the correlation between categorical and continuous morphological disparity. It remains to be seen if simulating categorical data under an evolutionary model affects the null trends that are derived.

MATERIAL AND METHOD

Generating trees

We followed a primarily simulation-based approach to characterizing the role of tree symmetry and branch length distribution in analyses of disparity (Smith *et al.* 2021). Data were simulated along six different generating trees (Fig. 1; Table 1). These trees were designed to form three paired contrasts, each of which varied in one aspect of tree shape. These were: (1) the symmetry of ultrametric trees; (2) the ‘stemminess’ of asymmetric, non-ultrametric trees; and (3) the ‘ultrametricity’ of asymmetric trees. For clarity, symmetry, sometimes referred to as balance, is discussed in terms of the extent to which nodes define subgroups of equal diversity. ‘Stemminess’ is used here to

FIG. 1. Generating trees used to simulate discrete character data. A, symmetrical. B, asymmetrical. C, long-leafed. D, long-stemmed. E, ultrametric. F, non-ultrametric. Trees presented are not to scale. See Table 1 for basic dimensions and further details.



refer to the proportion of the total branch length of a tree that subtends its nodes. The term ‘ultrametricity’ is used to refer to root-to-tip distance uniformity. Trees are only described as ultrametric when they have equal root-to-tip distances. For a graphical representation of these parameters of tree shape, see Figure 2. Branches that subtend internal nodes are referred to as internal or stem branches, while those that subtend tips (i.e. external nodes) are dubbed external or leaf branches. All six trees are composed of 32 tips and were fully bifurcating. For the constituent trees of each paired contrast, edge lengths were isometrically scaled to ensure equivalency in the sum of all branch lengths between them both. For brevity, the sum of all branch lengths of a tree is hereafter referred to as its total branch length.

Symmetry of ultrametric trees. Two generating trees with diametrically opposed topologies were employed: one fully symmetrical (Figs 1A, 3A), the other entirely asymmetrical (Fig. 1B). Both trees were ultrametric. For ease of reference, data simulated along these generating trees are hereafter described as symmetrical or asymmetrical (e.g. symmetrical matrices, asymmetrical matrices).

Stemminess of asymmetrical, non-ultrametric trees. The generating trees employed were: a tree with low stemminess (Figs 1C, 3B), where all internal branches were one tenth of the length of the external; and a tree with high stemminess (Fig. 1D), where the stem-leaf branch length ratio was reversed (i.e. where all internal branches were 10 times the length of the externals). Both trees were fully asymmetric. From this point onwards, these trees, and the data simulated along them, are dubbed ‘long-stemmed’ and ‘long-leafed’ respectively (e.g. long-leafed matrices, long-stemmed matrices).

Ultrametricity of asymmetric trees. The two generating trees employed were an ultrametric tree (Fig. 1E) and a non-ultrametric tree (Fig. 1F). Root-to-tip distances varied across all lineages of the non-ultrametric tree. Both trees were fully asymmetrical. For the sake of simplicity, data generated using these trees are described as either ultrametric or non-ultrametric (e.g. ultrametric matrices, non-ultrametric matrices).

Each paired contrast sought to isolate the signal of the tree shape parameter in question in the resulting perceptions of morphospace occupation. As such, all aspects of tree shape not under investigation were controlled. While such trees are generally considered to be unrealistic, here they were necessary for guaranteeing equivalency in all other phylogenetic structural properties to ensure that any differences recovered were a consequence of the shape parameter under investigation.

Simulation protocol

Data were simulated along the generating trees in R v3.6.3 (R Core Team 2020), employing the `rTraitDisc` function included in the `ape` R package (Paradis & Schliep 2019). Parallelization of simulations was achieved through employment of the parallel R package (R Core Team 2020). Discrete binary characters were simulated independently under an equal-rates model, so rates of change were equal across both possible character state transitions. Gamma distributions were generated for each generating tree by setting parameters *shape* to 0.44 and *scale* to the total branch length of the respective tree. We chose the value 0.44 for the *shape* parameter through trial and error to obtain the most visually realistic distributions. *Scale* was set to the branch length of the generating tree employed to ensure that the

TABLE 1. Dimensions and tree shape parameters of generating trees used to simulate discrete character data.

	Symmetrical	Asymmetrical	Ultrametric	Non-ultrametric	Long-leafed	Long-stemmed
Total branch length	38.44	38.44	26.35	26.35	23.24	23.24
Tree length	3.1	2.26	1.55	13.175	2.656	21.07
Internal branch length	0.62	0.073	0.05	0.425	0.066	0.7
External branch length	0.62	Variable	Variable	0.425	0.664	0.07
Colless index	0	465	465	465	465	465
Stemminess index	0.23	0.03	0.031	0.057	0.009	0.117

Branch lengths are in units of time.

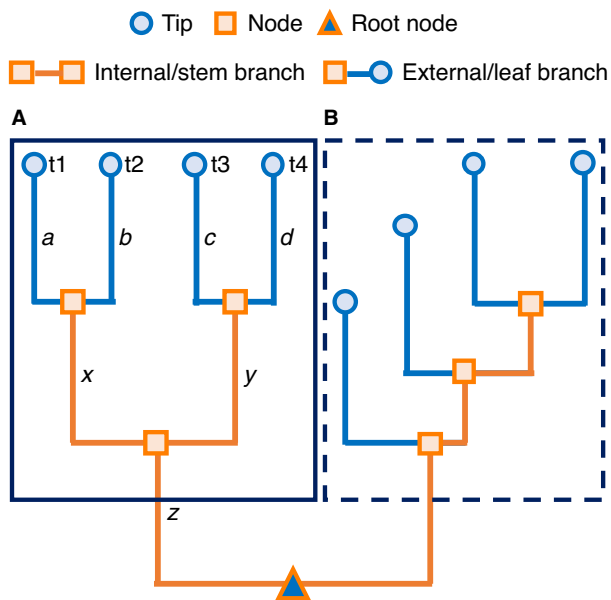


FIG. 2. Graphical depiction of the terminology used to describe tree shape. Subclades A and B contrast in both symmetry and ultrametricity. As all descendants of all nodes of subclade A are of equal diversity, its topology can be described as symmetrical or balanced. As this is not true for subclade B, its topology can be described as asymmetrical or imbalanced. The subtree of subclade A can be described as ultrametric as the root-to-tip distances of all its members are equal. The same cannot be said for the members of subclade B. Following Fiala & Sokal (1985), the stemminess of a tree is calculated as the proportion of total branch length located on the internal branches of a tree, averaged across all nested subclades. For example: the stemminess of group t1–t2 is calculated as $x/(a + b + x)$. The stemminess of group t3–t4 is $y/(c + d + y)$. The stemminess of subclade A (group t1–t4) is the mean of the stemminess values of its constituent subclades, t1–t2 and t3–t4. Note that in this scenario, the node uniting groups t1–t2 and t3–t4 is treated as the root node of the subclade, hence the branch subtending it is not included in the stemminess calculation.

distributions provided rates that facilitated the simulation of character data with consistency index (CI) values greater than 0.259, the lower limit of empirical data (Sanderson & Donoghue 1989). These combinations of values yielded

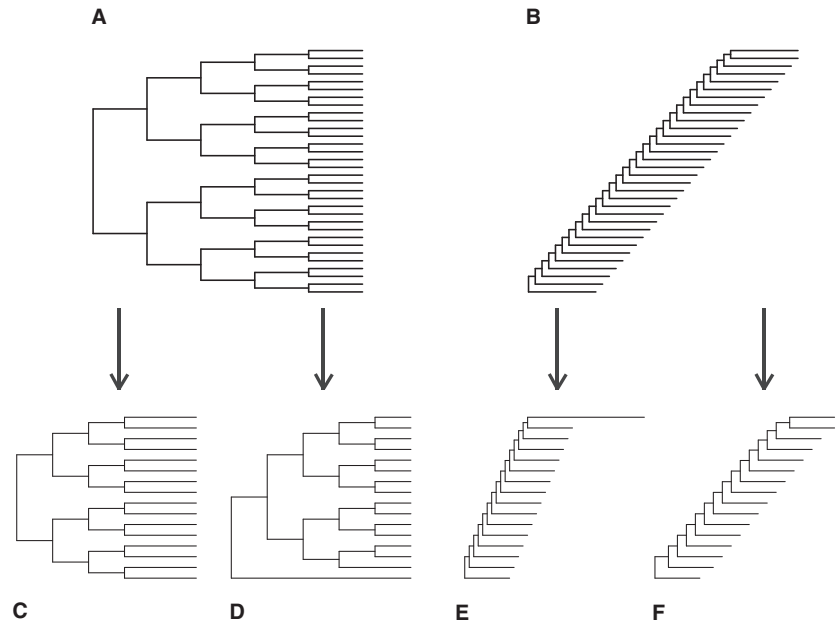
realistic distributions when plotted, relative to empirical data. For consistency within and between matrices sharing a generating tree, each gamma distribution was sampled 1000 times and the mean value calculated. This mean value was used as the rate of change for all character data simulated along the relevant generating tree.

To reflect the sampling one might expect of empirical cladistic datasets, autapomorphic characters were discarded and the simulation process repeated until only parsimony-informative sites remained. Simulated characters were concatenated to form 32, 65 and 260-character matrices to assess whether the taxon:character ratio of a dataset altered the impact of sampling. The CI was calculated for each of these matrices and those with unrealistically low CI values (i.e. <0.259) were discarded and replaced, ensuring that only matrices with approximately realistic distributions of homoplasy were analysed (see O'Reilly *et al.* 2016). In total, six datasets of 1000 matrices were generated, one for each generating tree. These datasets contained both node and tip states for each character. These node states are analogous to the output of ancestral state estimation. All other rTraitDisc arguments were left as their default settings.

Empirical datasets

Analyses of empirical data were conducted to see whether the trends we observed in our simulations are reflected in empirical data. We analysed four empirical datasets: a mammal dataset from Beck & Lee (2014); a crinoid dataset from Wright (2017); and two theropod datasets, one from Brusatte *et al.* (2014), the other from Bapst *et al.* (2016). These are respectively referred to as the mammal, crinoid, coelurosaur and avetheropod datasets herein (Smith *et al.* 2021). Each dataset included a time-scaled phylogeny, a morphological character matrix, and first and last occurrence dates for all included taxa. These four datasets were previously curated by Guillaume & Cooper (2018) and used to test the efficacy of various time-binning methodologies. Both the original and curated versions of the datasets are available for download with

FIG. 3. Subtrees derived through pruning of the symmetrical and long-leafed generating trees. A–B, original, 32-tip trees: A, symmetrical; B, long-leafed. C–F, pruned, 16-tip trees: C, symmetrical; D, asymmetrical; E, long-leafed; F, long-stemmed. Trees are not to scale.



their accompanying papers. The curated datasets of Guillerme & Cooper (2018) were employed in the analyses described below (see Table 2 for dataset details). Prior to analysis, each dataset was pruned of taxa only present in the matrix but not the tree and vice versa. After pruning, the taxa with the shortest and longest root-to-tip distances were designated as endmembers for each dataset. These taxa were retained in all subsamples analysed to ensure the same phylogenetic bracket was represented in each iteration. A small number of outgroup taxa in the matrices of coelurosaur, mammal and avetheropod branched off prior to their respective lower endmembers; these taxa were also pruned out of the datasets prior to taxic disparity analysis (Table 2). We adopted this bracketing procedure in an effort to ensure comparability between subtrees in their rooting and temporal range.

Distance matrix computation

Pairwise distance matrices of $n \times n$ dimensions, where n represents the number of operational taxonomic units (OTUs) present in the source data, were constructed for each matrix using the `MorphDistMatrix` function included in the `Claddis` R package (Lloyd 2016; renamed ‘`calculate_morphological_distances`’ from v0.6.1 onwards). The $n \times n$ distances were calculated using the Gower dissimilarity coefficient (Gower 1971) for both the simulated and empirical datasets to permit direct comparison and accommodate missing data present in the latter (Anderson & Friedman 2012). There is a limit to the ability of the Gower coefficient to handle missing data; pairwise distances cannot be calculated if pairs of taxa lack

overlapping data. While the simulated, crinoid and mammal datasets all returned complete distance matrices, the coelurosaur and avetheropod datasets did not. To address this issue, the taxa responsible for these gaps were identified and trimmed out of the datasets using the `Claddis` `TrimMorphDistMatrix` function (Table 2; renamed to ‘`trim_matrix`’ from v0.6.1 onwards). To this end, 9 and 19 taxa were removed from the avetheropod and coelurosaur datasets respectively.

Ordination

Distance matrices were transformed through classical multidimensional scaling (MDS), also referred to as principal coordinates analyses (PCO/PCoA), using the `Claddis.ordination` function included in the `disPRity` R package (Guillerme 2018). Prior to ordination, Gower coefficients were first transformed through application of the Cailliez correction (Cailliez 1983), to address the potential issue of negative eigenvalues, and a square root term, to make the distances more Euclidean (Wills 2001).

Characterizing and describing morphological disparity

Disparity was characterized using four commonly employed indices: pre-ordination mean pairwise distance; and post-ordination sum of variances, sum of ranges, and mean distance from centroid. These indices were not chosen for their independence or ability to characterize different aspects of morphospace; indeed, as mean distance from centroid and sum of variances are deterministically

TABLE 2. Details of empirical datasets analysed herein, modified from Guillaume & Cooper (2018).

	Mammal	Coelurosaur	Avetheropod	Crinoid
Source	Beck & Lee (2014)	Brusatte <i>et al.</i> (2014)	Bapst <i>et al.</i> (2016)	Wright (2017)
No. characters	402	853	374	87
No. taxa (original)	102	140	89	42
Age range (Ma)	171.8–0	168.5–66	207.2–66	485.4–373.2
Missing data percentage	17.93%	71.27%	51.54%	50.82%
Inapplicable data percentage	0%	0%	0.24%	0.06%
No. taxa removed to enable complete distance matrix	0	19	9	0
Endmember taxa (for selective subsampling analysis)	<i>Juramaia</i> , <i>Solenodon</i>	<i>Zuolong</i> , <i>Rinchenia</i> <i>mongoliensis</i>	<i>Sinraptor</i> , <i>Zanabazar junior</i>	<i>Apektocrinus</i> , <i>Amabilicrinus</i>
No. taxa outside endmember bracket removed (for selective subsampling analysis)	3	2	1	0

related, their covariance is to be expected. Rather, they were chosen for their pervasiveness in empirical research to ensure our results could be employed in the interpretation of as great a variety of contexts as possible. The post-ordination indices were computed using the *dispRity* function included in the *dispRity* R package (Guillaume 2018). Here we describe morphological disparity in terms of morphospace occupation. The structure of a morphospace has two key components: the amount of space occupied, and the spacing of points within the occupied region. Hereinafter, these components are respectively referred to as the size and density of the distribution within morphospace of a clade.

In our analyses, we employ the mean pairwise distance to characterize the density with which taxa occupy a morphospace and the sum of ranges as a measure of the size of occupied area. We also employ the sum of variances and mean distance from centroid to characterize morphospace. Interpretations of these indices in the context of disparity analyses are varied. Foote (1992) described variance as ‘a measure of morphological dispersion’ but did not provide further detail on what is meant by this. He did, however, employ ‘total variance’ as a measure of ‘morphological variety’, which he defined as ‘the variance in form or the amount of morphological space occupied’ (Foote 1992). Wills’ (2001) interpretation of the sum of variances of a clade incorporates both the size of the space occupied and spacing of the points within; points with a low sum of variances will cluster together, whereas taxa with a higher value will scatter across morphospace or congregate near the edges of the occupied area. Other interpretations have considered the sum of variances as a strict measure of the amount of morphospace occupied (Guillaume *et al.* 2020b) or the density with which taxa occupy an explored region (Hopkins & Gerber 2017). Wills’ (2001) interpretation is the most intuitive: the sum of variances incorporates aspects of size and density in

how it characterizes morphospace. However, as alluded to by Wills (2001), the characterization of density by the sum of variances breaks down when variance is high; it is conceivable that two groups of taxa, one scattered across morphospace and the other defined by clusters congregating at the limits of the occupied area, could present the same levels of variance, despite the markedly different patterns of intertaxon spacing. Put another way, the correlation of variance with density is contingent on a morphospace being normally distributed. The index is too coarse to informatively characterize the density of taxa within morphospace unless variance is low and the space normally distributed. As a measure of the amount of morphospace occupied, it is more reliable; the characterization of the size of the area occupied does not lose meaning when variance is high. As such, we use the sum of variances to characterize the size of the area occupied by taxa but recognize that at some levels, it also characterizes the density with which taxa populate the occupied region. We employ the mean distance from centroid as a measure of the size of the area occupied, following the same logic.

Analysis of simulated datasets

To evaluate the effect of sampling on taxic morphospace occupation, comparisons were made between the raw disparity distributions of the datasets contrasting in symmetry and stemminess for each combination of character number and index. Comparisons were also made between the datasets contrasting ultrametricity (see Appendix S1). However, these are not presented below as the branch lengths of the ultrametric and non-ultrametric trees incorporate a dating component that the symmetrical, asymmetrical, long-leafed and long-stemmed do not. As such, comparisons between the datasets contrasting in

ultrametricity are only comparable to empirical approaches within frameworks that account for time, such as disparity-through-time analyses. These analyses were conducted and are described in greater detail below.

The symmetrical (Figs 1A, 3A) and long-leafed trees (Figures 1C, 3B) were then selectively pruned to determine whether biases represented by variations in symmetry and stemminess could be introduced through subsampling. Two rounds of pruning were conducted: the first intended to produce a tree that preserved the shape of the original, while the second aimed to do the opposite. This approach produced two 16-tip trees for each 32-tip original; one pair contrasting in symmetry (Fig. 3C, D), the other in stemminess (Fig. 3E, F). Using the new 16-tip trees for reference, the symmetrical and long-leafed datasets were then subsampled to match. The result was four new 16-tip datasets comprised of a tree and 1000 matrices. The morphological disparity of these new matrices was characterized, and the resulting distributions compared to those of the original trees.

Disparity-through-time analyses were conducted to determine whether sampling altered perceptions of evolving morphospace occupation. Relative occurrence times were assigned to the tips and nodes of each generating tree; time '0' was assigned to the root node, while time '1' was assigned to the tip furthest from the root. Each generating tree was then partitioned into six equal-sized time partitions, where each partition spanned one sixth of the overall length of the tree. Tips and nodes sampled within each partition were recorded, as were any tips with subtending branches that crossed the upper partition limit. We refer to such tips as bisected lineages going forward. The first 100 simulated matrices of each generating tree were partitioned into six subsets matching the composition of the partitions. Tips, nodes and bisected lineages were analysed in the following combinations to evaluate their relative contributions to perceptions of disparity-through-time: tips only; tips and nodes; tips and bisected lineages; tips, nodes and bisected lineages. Each partition was bootstrapped and rarefied prior to disparity index calculation to ensure equivalency in diversity between partitions. Each binned subsample of each matrix was randomly subsampled with replacement 100 times, with the size of the new subsamples set to three. Disparity index values were then calculated and their distributions through time compared for each generating tree.

All analyses of simulated data were conducted in R by employing a range of functions included in the stats v3.6.2, ape v5.3 (Paradis & Schliep 2019), phangorn v2.5.5 (Schliep 2011), dispRity v1.4.1 (Guillerme 2018), Claddis v0.3.4 (Lloyd 2016), phytools v0.7-20 (Revell 2012) and apTreeshape v1.5-0 (Bortolussi *et al.* 2006) packages.

Analysis of empirical datasets

To validate the results of the simulations, each empirical dataset was subsampled to create contrasts in tree symmetry and stemminess. This was achieved by first selectively pruning each empirical tree four times over to emphasize specific aspects of tree shape (e.g. greater asymmetry, low stemminess). This pruning yielded four subsampled trees for each empirical dataset, each of which contained 50% of the tips of the original. To ensure these subsamples contrasted sufficiently, their shape was quantified. Tree symmetry was characterized using the Colless Index (Colless 1982), which measures the balance of a bifurcating tree by returning the sum of the absolute differences in number of descendants between each pair of lineages descending from each node. Stemminess was quantified using the homonymous index of Fiala & Sokal (1985), which is defined as the proportion of the total branch length of each sub-clade that is accounted for by the length of the subtending edge, averaged across all nodes of the tree. Using these trees for reference, the empirical matrices were then subsampled to match. The morphospace occupation of these matrices was then characterized, and the resultant disparity index values scaled to account for differences in total branch length between the subsampled trees and the original. Finally, the deviations of these scaled disparity index values from those characterizing the morphospace occupation of the original datasets were calculated.

In addition to the selective subsamples, 1000 random subsamples of the same size were drawn from each empirical matrix without replacement to facilitate the creation of a null distribution of disparity under subsampling. For each subsample, morphospace occupation was characterized and a complimentary tree derived through pruning of the original to match. Disparity index values were then scaled to compensate for differences in total branch length between each random subsample and the original tree. Finally, the deviations in disparity of each random subsample from that of the original dataset were calculated. These deviations made up the null distribution to which the deviations of the selective subsamples were compared to determine whether manifestations of tree shape are discernible from random variation. To quantify the significance of deviations, a p-value was derived for each selective subsample: for subsamples yielding greater disparity than most of the null distribution, p was calculated as $n/1000$, where n equals the number of random subsamples deviating more positively than the selective subsample; for selective subsamples with low disparity relative to the null distribution, p was calculated as $1 - (n/1000)$. An arbitrary threshold of $p = 0.05$ was employed when assessing the significance of the deviations.

Disparity-through-time analyses of the empirical datasets were also conducted. Each dataset partitioned by

stratigraphic age. Partitions containing too few taxa for disparity calculation (<3 taxa) were removed from the analysis. For each remaining partition, a matching subtree was generated through pruning of the original tree. The shapes of these subtrees were then quantified in terms of symmetry and stemminess. Subtree symmetry was measured using the Colless Index (Colless 1982). As the Colless Index has a non-linear relationship with diversity (Rodgers 1993), the value calculated for each time partition was divided by the maximum possible value for the number of taxa sampled. This gave a normalized Colless Index with a scale of 0–1, where 1 was indicative of a maximally asymmetrical topology and 0 of a maximally symmetrical. Subtree stemminess was measured with the homonymous index of stemminess defined by Fiala & Sokal (1985).

Each partition was bootstrapped and rarefied through random subsampling with replacement 100 times to ensure that perceived trends in disparity-through-time were not a product of discrepancies in taxonomic diversity. The size of these subsamples was set as the number of taxa included in the least diverse partition of the series. The morphospace occupation of each of these subsamples was then characterized using mean pairwise distance, sum of variances, mean distance from centroid, and sum of ranges. For time partition, median values were identified, and 90% confidence intervals calculated.

Finally, changes in disparity-, symmetry- and stemminess-through-time were assessed for positive covariance. For each combination of empirical dataset, disparity index, and tree shape index, the number of intervals (i.e. changes in disparity and tree shape from one time partition to the next) displaying positive covariance was counted and the overall percentage displaying this relationship calculated. Intervals were excluded from these analyses if tree shape remained unchanged from one time partition to the next, or if one of the partitions defining an interval was too species-poor to be ordinated (this was only relevant in analyses employing post-ordination indices of disparity).

All analyses of the empirical datasets were conducted using the same R packages as in the simulations.

Plotting

All plots were generated in R using functions included in the `ggplot2` package family (Wickham 2016).

RESULTS

In many of the analyses conducted, an increase in the size of the area of morphospace occupied by a clade was

accompanied by a decrease in density. For the sake of brevity, such concomitant changes in morphospace occupation size and density are discussed as changes in dispersion going forward; greater dispersion translates to an increase in space occupied coupled with a decrease in density, and vice versa.

Simulated data

In all analyses of simulated data, the relationships recovered between morphospace occupation and tree shape did not change with character number. While the variance in disparity index values decreased as character number increased, the overall trends remained the same. As such, only the results of the analyses of the 260-character matrices are presented and discussed (see Appendix S1 for 32-character and 65-character results). Additionally, as all four indices of disparity covaried in the disparity-through-time analyses conducted, only the mean pairwise distance results are figured. However, the descriptions below reflect the behaviour of all indices.

Topological symmetry. Mean pairwise distance, sum of variances, and mean distance from centroid exhibited a positive relationship with tree symmetry (Fig. 4A–C). A negative relationship was recovered between symmetry and sum of ranges (Fig. 4D). In terms of morphospace occupation, symmetry correlated with greater overall dispersion, but the asymmetrical tree yielded the greater divergence in peripheral morphotypes.

Regardless of the change affected, subsampling of the simulated matrices to manipulate the balance of the symmetrical generating tree greatly increased disparity index variance (Fig. 5). Consequently, there was considerable overlap between the topologically contrasting subsamples. Overall, emphasizing symmetry yielded subsamples with greater morphospace dispersion than pruning for asymmetry. The original and subsample distributions of mean pairwise distance, sum of variances, and mean distance from centroid were comparable in their median values (Fig. 5A–C). Most subsamples exhibited much lower sum of ranges than the originals (Fig. 5D). In other words, the original matrices characterized a greater range of morphologies, but were comparable in their morphospaces to those of the subsamples in terms of dispersion.

Disparity-through-time analyses of the symmetrical and asymmetrical matrices yielded distinct temporal patterns. As the analysed trees were both ultrametric, all tips were concentrated in the final partition. Consequently, informative comparisons between the disparity-through-time trends of the two trees could only be made when nodes were included in the time partitions, hence these are the only results figured (Fig. 6). When tips and nodes were

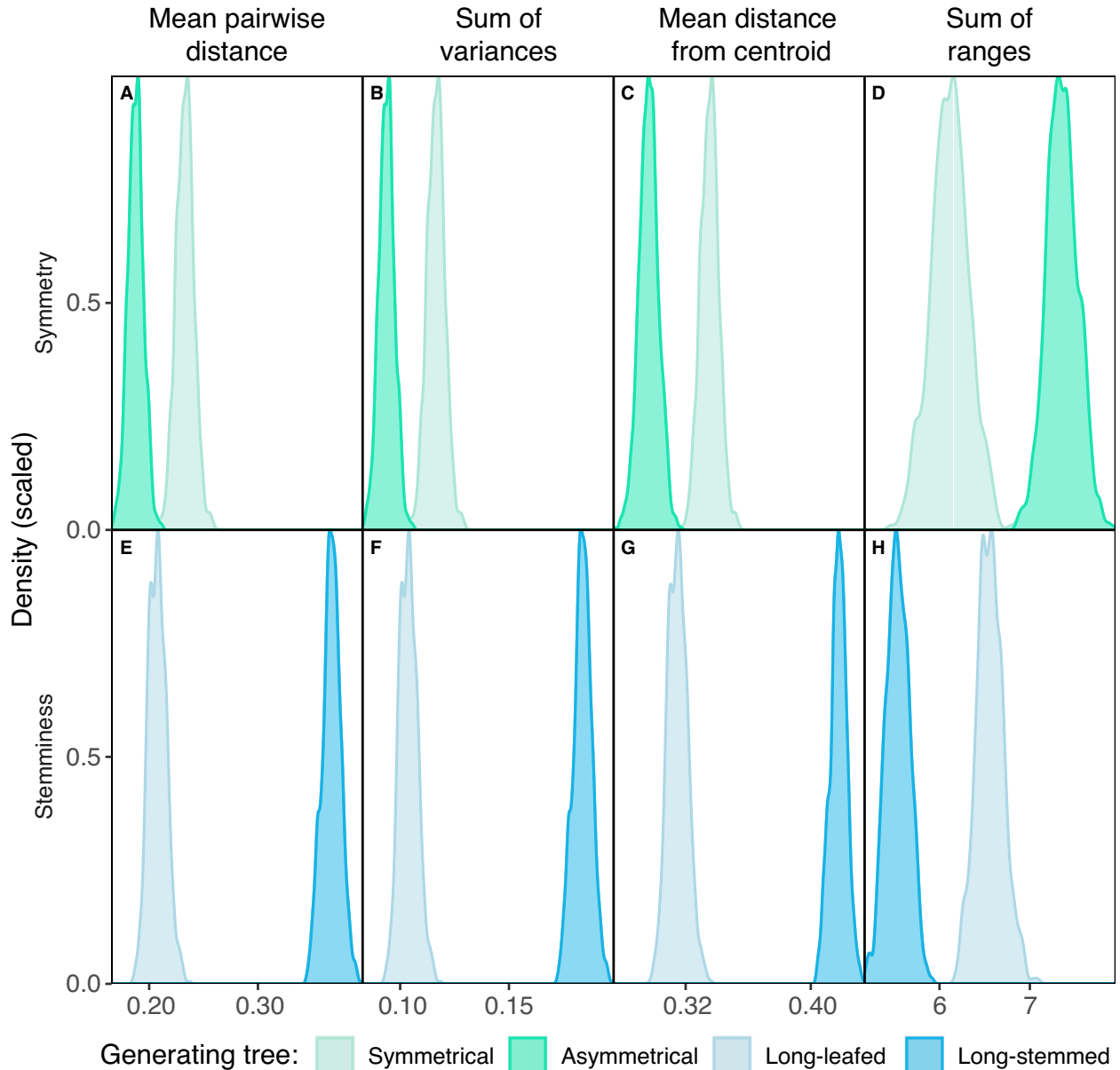


FIG. 4. Mean pairwise distance (A, E), sum of variances (B, F), mean distance from centroid (C, G), and sum of ranges (D, H) of matrices simulated along the symmetrical, asymmetrical, long-leaved and long-stemmed generating trees. Each distribution represents the disparity index values of 1000 matrices.

analysed collectively, morphospace dispersion increased gradually along the symmetrical tree, whereas across the asymmetrical tree it remained low until the evolution of the tips (Fig. 6A). The symmetrical matrices consistently exhibited greater dispersion within morphospace than the asymmetrical, although the difference between the tips of the two was marginal. When bisected lineages were incorporated, both trees displayed gradual increases in morphospace dispersion through time, although the trend was much weaker along the asymmetrical tree than the symmetrical (Fig. 6B). Between partitions 1 and 3, the

morphospace occupation of the asymmetrical matrices was more dispersed than the symmetrical. From partitions 4 to 6, the dispersion exhibited by the symmetrical matrices exceeded that of the asymmetrical.

Stemminess. The long-stemmed matrices exhibited greater mean pairwise distances, sums of variances, and mean distances from centroid than their long-leaved counterparts (Fig. 4E–G). An inversion of this relationship was recovered between tree stemminess and sum of ranges (Fig. 4H). Collectively, these results indicate that

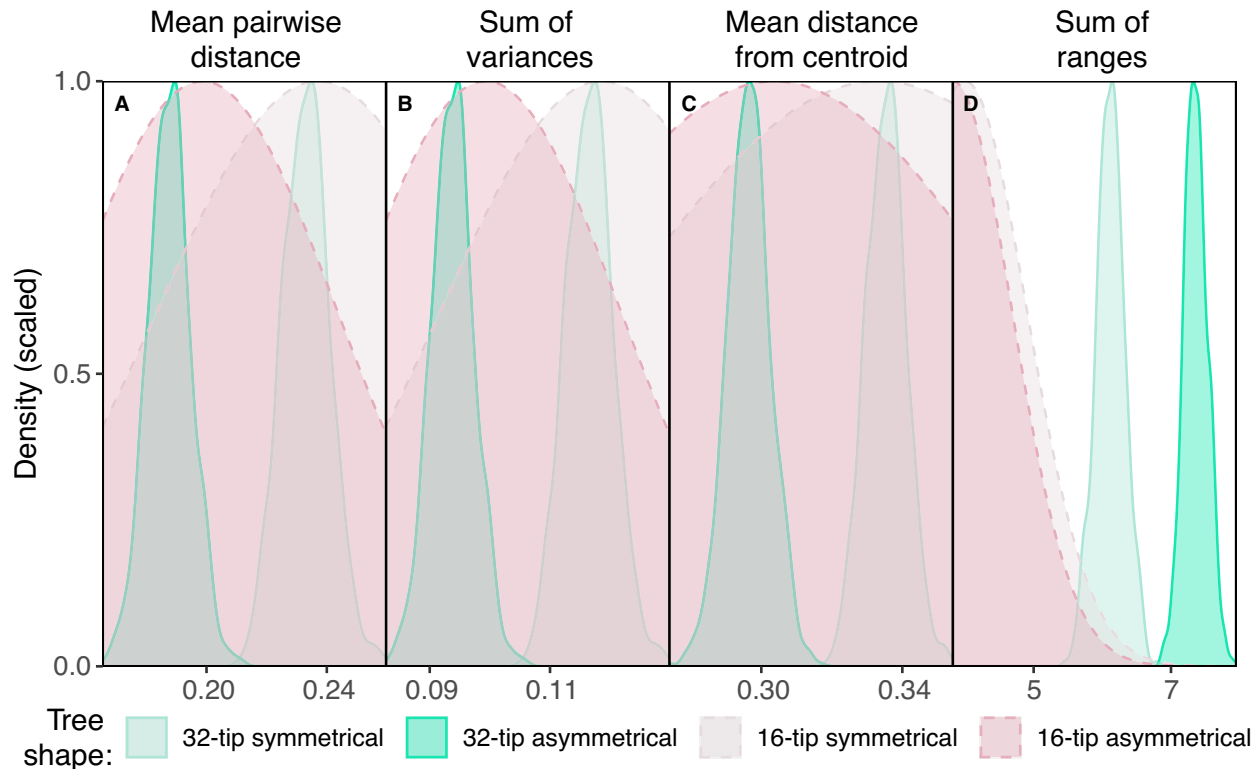


FIG. 5. Comparison of matrices simulated along trees contrasting in symmetry and subsamples derived through manual pruning in terms of disparity. The 16-tip subtrees and their corresponding matrices were derived through the subsampling of the 32-tip symmetrical tree and the matrices originally simulated along it. Each distribution represents the disparity index values of 1000 matrices.

stemminess correlates with dispersion, while long external branches predict a greater phenotypic range.

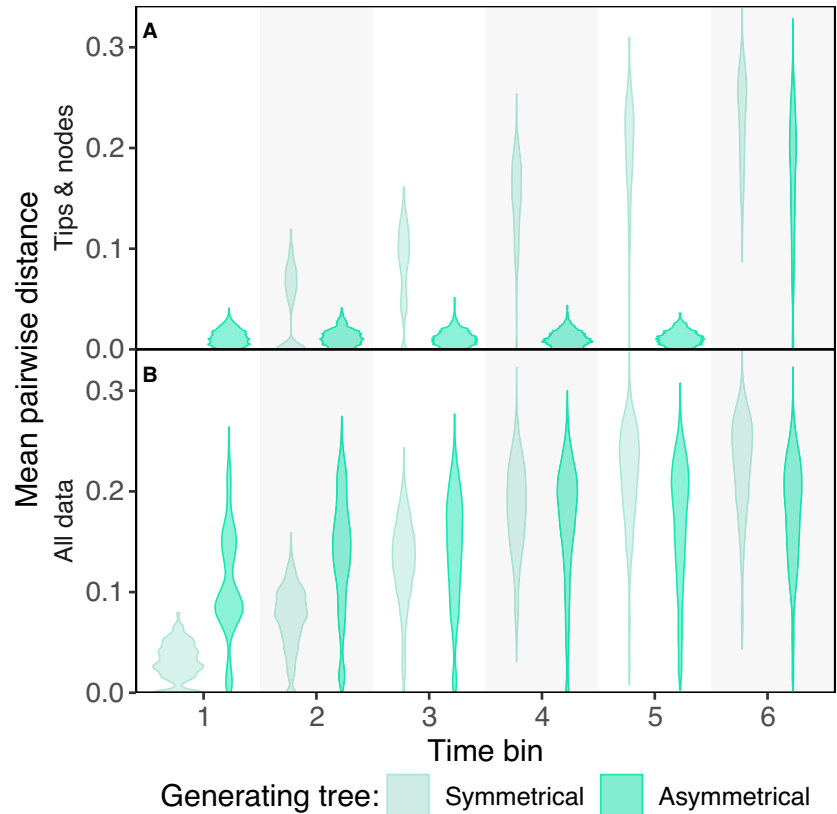
Subsampling of the long-leaved matrices to manipulate the stemminess of the underlying tree produced disparity index distributions that largely overlapped, regardless of index employed or contrast in branch length distribution imposed (Fig. 7). The variance of the subsample distributions exceeded that of the originals. Emphasizing tree stemminess through subsampling yielded lower sums of ranges but greater mean pairwise distances, sums of variances, and mean distances from centroid than minimizing it (Fig. 7A–D). The effect size of stemminess on morphospace occupation was much smaller in the subsamples than it was in the original datasets. However, the polarity of the relationship was the same; stemminess correlated with greater dispersion in morphospace, and long external branches were predictive of greater phenotypic range. Overall, the sum of variances, mean pairwise distance and mean distance from centroid of the subsamples did not deviate significantly from that of the original long-leaved matrices they were derived from (Fig. 7A–C), indicating that the datasets exhibited comparable distributions in morphospace. However, a small number of subsamples exhibited mean distances from centroid comparable to those of the original long-stemmed matrices. There was

almost no overlap in the sums of ranges of the subsampled and original matrices (Fig. 7D); a handful of subsamples exhibited comparable values to those of the original long-stemmed matrices.

Disparity-through-time analyses of the long-leaved matrices using tips alone recovered a mostly constant pattern of morphospace occupation along the tree, with the exception of a small increase in dispersion between partitions 2 and 3 (Fig. 8A). In contrast, the long-stemmed matrices exhibited stasis from partition 1 to 5 before slightly increasing in morphospace dispersion in the final partition. Generally, the long-leaved matrices exhibited greater dispersion than the long-stemmed. When tips and bisected lineages were analysed together, the long-leaved tree exhibited an increase in morphospace dispersion between partitions 1 and 2, followed by stasis from partition 2 to 5, and a marked decrease in the final partition (Fig. 8B). In contrast, morphospace occupation remained static through time along the long-stemmed tree. The long-leaved matrices were consistently more dispersed within morphospace than the long-stemmed.

Using tips and nodes to compute disparity-through-time along the long-leaved tree yielded an increase in morphospace dispersion from partition 1 to 3, followed by a gentle decrease across the remaining partitions

FIG. 6. Mean pairwise distance-through-time of the matrices simulated along the symmetrical and asymmetrical generating trees, where partition 1 includes the root and partition 6 the tips. ‘Tips & nodes’ (A) and ‘all data’ (B) describe the composition of the time partitions. The symmetrical and asymmetrical trees were of unequal tree length, hence there was some variation in time partition duration between the two. However, all partitions along each tree were of equal size. Violins capture the complete distribution of 10 000 disparity index values (100 bootstraps of 100 matrices).



(Fig. 8C). Along the long-stemmed tree, morphospace occupation was static through time. With the exception of the first partition, where the morphospace occupation of the long-leafed matrices was much more compact than the long-stemmed, the differences between the disparity-through-time trends of the two trees were minor; generally, the long-leafed matrices exhibited somewhat greater dispersion than the long-stemmed. When tips, bisected lineages and nodes were analysed collectively, dispersion increased between partitions 1 and 3, plateaued from 3 to 5, and then decreased marginally in the final partition along the long-leafed tree (Fig. 8D). Morphospace occupation remained largely constant along the long-stemmed tree, although a slight decrease in dispersion through time was evident. In partition 1, the long-stemmed matrices generally exhibited greater dispersion within morphospace than the long-leafed. Both sets of matrices were comparable in their average morphospace occupation in partition 2. However, between partitions 3 and 5, the dispersion of the long-leafed matrices surpassed that of the long-stemmed overall. In the final partition, both datasets were once again comparable in their average morphospace occupation, although some long-stemmed matrices exhibited distributions in multidimensional space that were much more compact than the most tightly clustered long-leafed comparators.

Ultrametricity. Owing to an absence of data in partitions 1–4, the tips of the ultrametric and non-ultrametric trees did not yield informative disparity-through-time trends, and so were not figured (Fig. 9). Disparity-through-time analysis using both tips and bisected lineages yielded incremental increases in the morphospace dispersion of the most disparate subsamples along both trees from partitions 4–6 (Fig. 9A). However, the majority of matrices were consistent in their morphospace occupation through time, with the ultrametric generally exhibiting greater dispersion than the non-ultrametric. Including tips and nodes in disparity index calculation produced different trends. Along the ultrametric tree, the size of the area occupied remained small and densely packed until the inclusion of tips, where it expanded and became sparser (Fig. 9B). In contrast, morphospace occupation along the non-ultrametric tree was static through time. The non-ultrametric matrices displayed greater dispersion within morphospace between partitions 1 and 5 than did their ultrametric counterparts, with the polarity of the relationship flipping in partition 6. When tips, nodes, and bisected lineages were analysed collectively, morphospace dispersion increased incrementally between partitions 1 and 4 before plateauing along the ultrametric tree (Fig. 9C). The non-ultrametric tree yielded a static disparity-through-time trend across all partitions. The

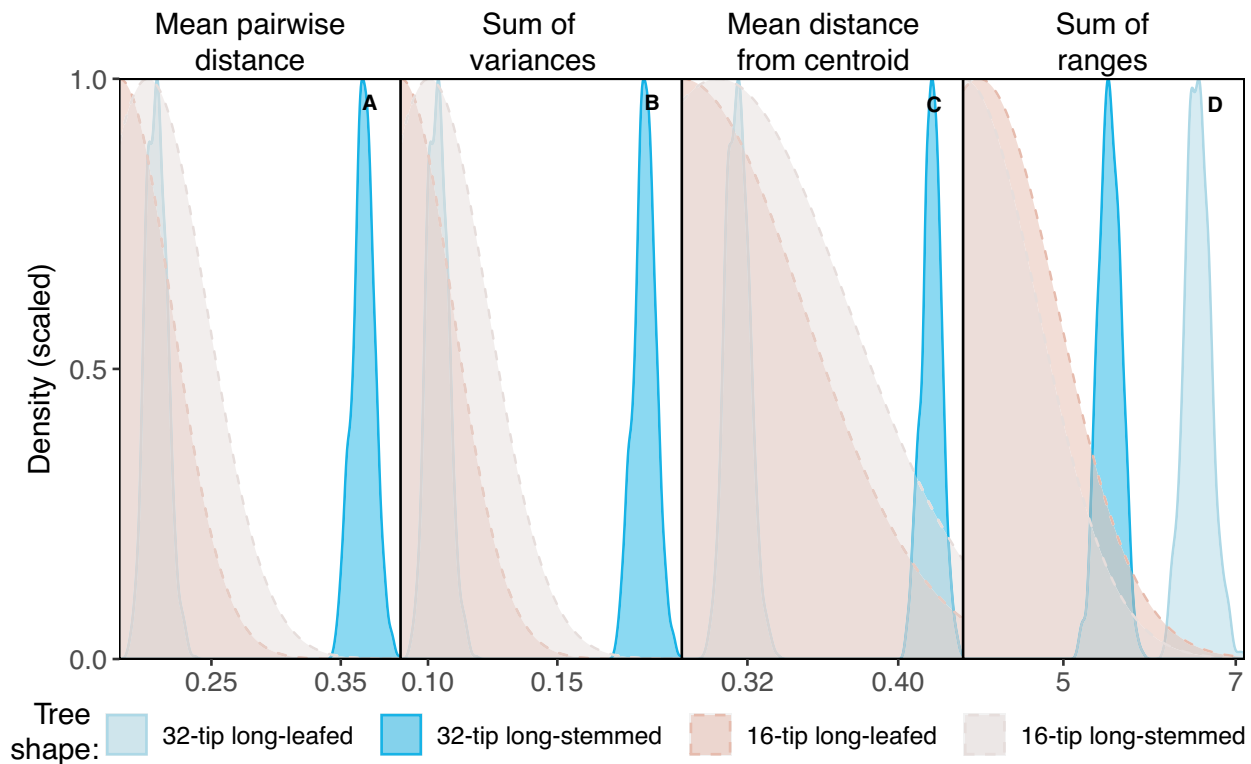


FIG. 7. Comparison of matrices simulated along trees contrasting in stemminess and subsamples derived through manual pruning in terms of disparity. The 16-tip subtrees and their corresponding matrices were derived through the subsampling of the 32-tip long-leafed tree and the matrices originally simulated along it. Each distribution represents the disparity index values of 1000 matrices.

ultrametric matrices were consistently more disparate than the non-ultrametric, with the differences between the two increasing with time.

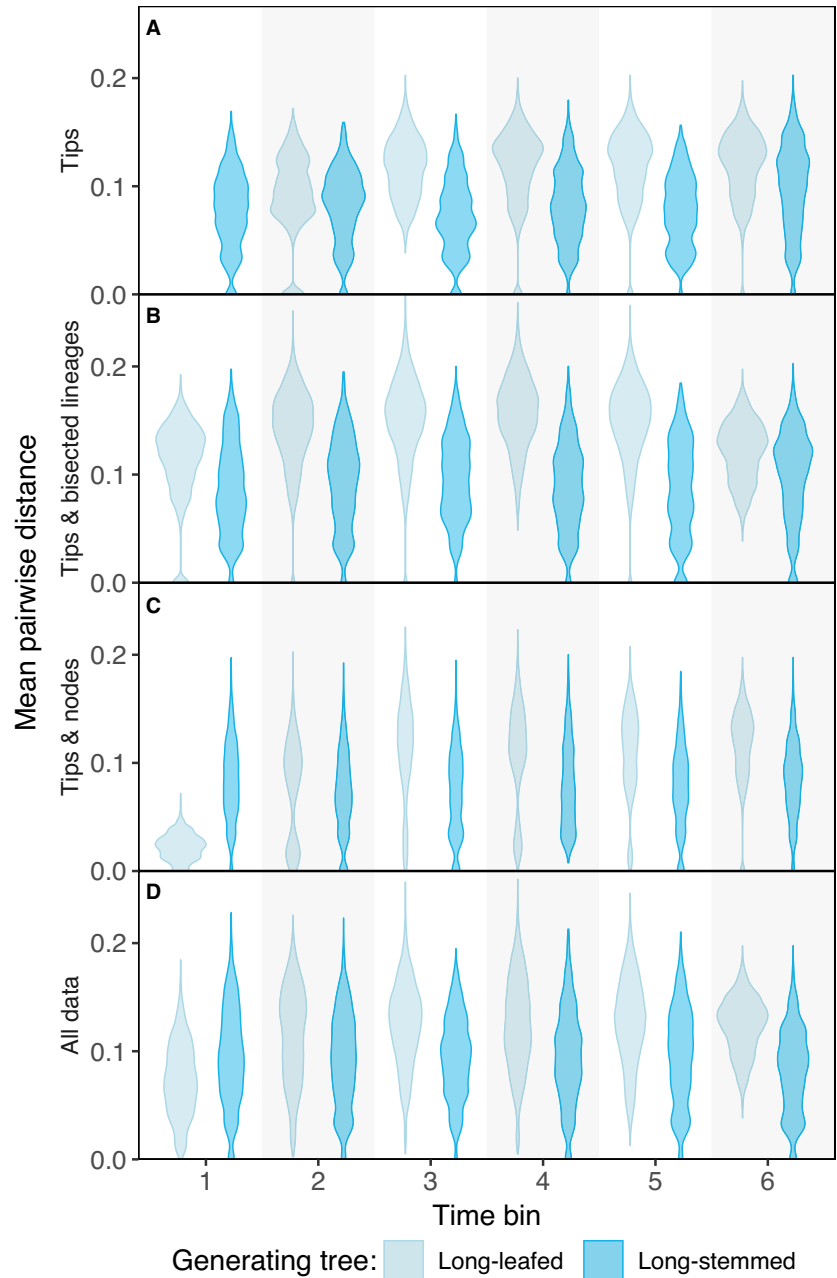
Empirical data

Symmetry. The coelurosaur dataset exhibited significantly greater dispersion within morphospace across all indices when subsampled to maximize tree symmetry than would be expected under random sampling (Fig. 10; Table 3). Subsampling to emphasize topological asymmetry produced a more complex pattern of morphospace occupation: the mean pairwise distance of the subsample was significantly lower ($p = 0.024$); the sum of variances significantly higher ($p = 0.001$), and sum of ranges and mean distance from centroid indistinguishable from that of random subsamples. The mean pairwise distance ($p = 0.038$) and sum of variances ($p = 0.005$) of the symmetrical subsample of the avetheropod dataset was significantly higher than would be expected through random subsampling. The sum of ranges and mean distance from centroid values were indistinguishable from those of random subsamples (Table 3). When topological asymmetry was emphasized, subsampling of the avetheropod dataset

yielded a distribution within morphospace indistinguishable from that of a random subsample across all indices. In terms of morphospace occupation, subsampling to manipulate topological symmetry did little to differentiate the resulting subsets of the mammal dataset from random subsamples (Table 3). Only the sum of variances of the symmetrical subsample deviated significantly, surpassing that of most random subsamples ($p = 0.001$). The symmetrical subsample of the crinoid dataset exhibited significantly higher sum of variances ($p = 0.028$), significantly lower mean distance from centroid ($p = 0.048$), and unremarkable mean pairwise distance and sum of ranges (Table 3). Conversely, emphasizing asymmetry through subsampling also yielded significantly higher sum of variances ($p = 0.009$). The mean pairwise distance, mean distance from centroid, and sum of variances of the crinoid asymmetrical subsample were not distinguishable from those of random subsamples.

In the disparity-through-time analyses, tree symmetry covaried with morphospace dispersion across a majority of intervals for all four empirical datasets (Fig. 11; Table 4), regardless of how the amount of morphospace occupied was characterized. However, in the case of the avetheropod dataset, the percentage of intervals displaying positive covariance between dispersion and symmetry was

FIG. 8. Mean-pairwise-distance-through-time of the matrices simulated along the long-leafed and long-stemmed generating trees, where partition 1 includes the root and partition 6 the tips. ‘Tips’ (A), ‘tips & bisected lineages’ (B), ‘tips & nodes’ (C) and ‘all data’ (D) describe the composition of the time partitions. The long-leafed and long-stemmed trees were of unequal tree length, hence there was some variation in time partition duration between the two. However, all partitions along each generating tree were of equal size. Violins capture the complete distribution of 10 000 disparity index values (100 bootstraps of 100 matrices).



only just above 50% when the size of the area occupied was captured using sum of variances and sum of ranges.

Stemminess. Maximizing the stemminess of the coelurosaur dataset through subsampling resulted in significantly greater dispersion within morphospace, regardless of which index was used to characterize the size of the area occupied (Fig. 4; Table 3). Conversely, subsampling to minimize stemminess also produced a subset with significantly higher sum of variances than the majority of random subsamples ($p = 0.03$). The

mean pairwise distance, mean distance from centroid, and sum of ranges of this same subsample of the coelurosaur dataset were all unremarkable. Stemminess positively correlated with dispersion within morphospace across all subsampling analyses of the avetheropod and crinoid datasets, with each subsample deviating significantly (Table 3). In analyses of the mammal dataset, stemminess positively correlated with density (mean pairwise distance) within morphospace. However, the size of the area of morphospace occupied by these subsets was not significantly different

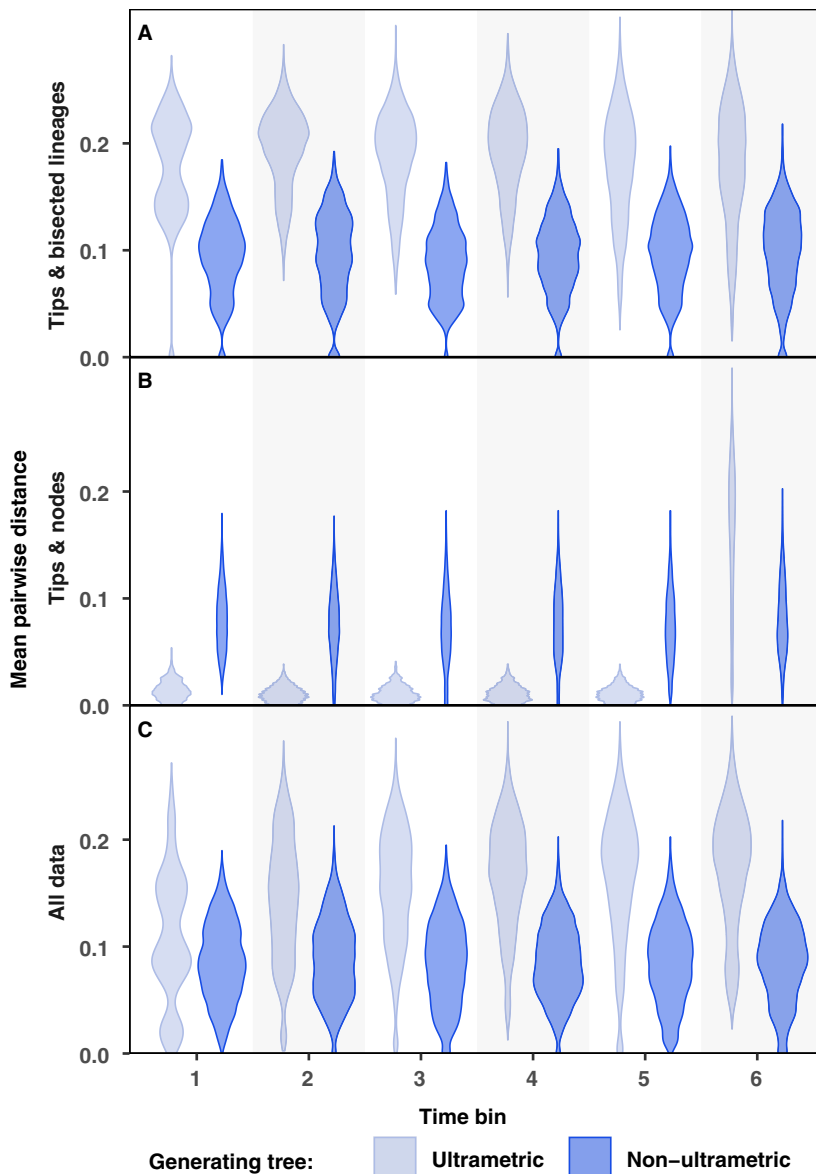


FIG. 9. Mean-pairwise-distance-through-time of the matrices simulated along the ultrametric and non-ultrametric generating trees, where partition 1 includes the root and partition 6 the tips. ‘Tips & bisected lineages’ (A), ‘tips & nodes’ (B) and ‘all data’ (C) describe the composition of the time partitions. The ultrametric and non-ultrametric trees were of unequal tree length, hence there was some variation in time partition duration between the two. However, all partitions along each generating tree were of equal size. Violins capture the complete distribution of 10 000 disparity index values (100 bootstraps of 100 matrices).

from that expected of random sampling, regardless of whether stemminess was maximized or minimized (Table 3).

Disparity-through-time analyses of the four datasets only recovered positive covariance between tree stemminess and morphospace occupation across a majority of time intervals when mean pairwise distance was employed to characterize the mammal dataset (Table 5). No such pattern of covariance was identified between the indices of size and stemminess for the mammal dataset, nor was one evident when the coelurosaur dataset was analysed. In analyses of the avetheropod and crinoid datasets, dispersion within morphospace negatively covaried across a majority of intervals, regardless of how morphospace occupation was characterized.

DISCUSSION

Simulated data analysis

Symmetry. The simulation results indicate that tree symmetry predicts the impact of sampling on morphospace. The more symmetrical a tree, the more dispersed we may expect a given clade to be in multidimensional space for a given model of character evolution. This relationship was replicable through subsampling; matrices generated using the same tree but differing in subsampled topology exhibited the same relationship with symmetry as those simulated along different trees, both in terms of polarity and magnitude. The overall dispersion of the subsampled matrices within morphospace was comparable to that of

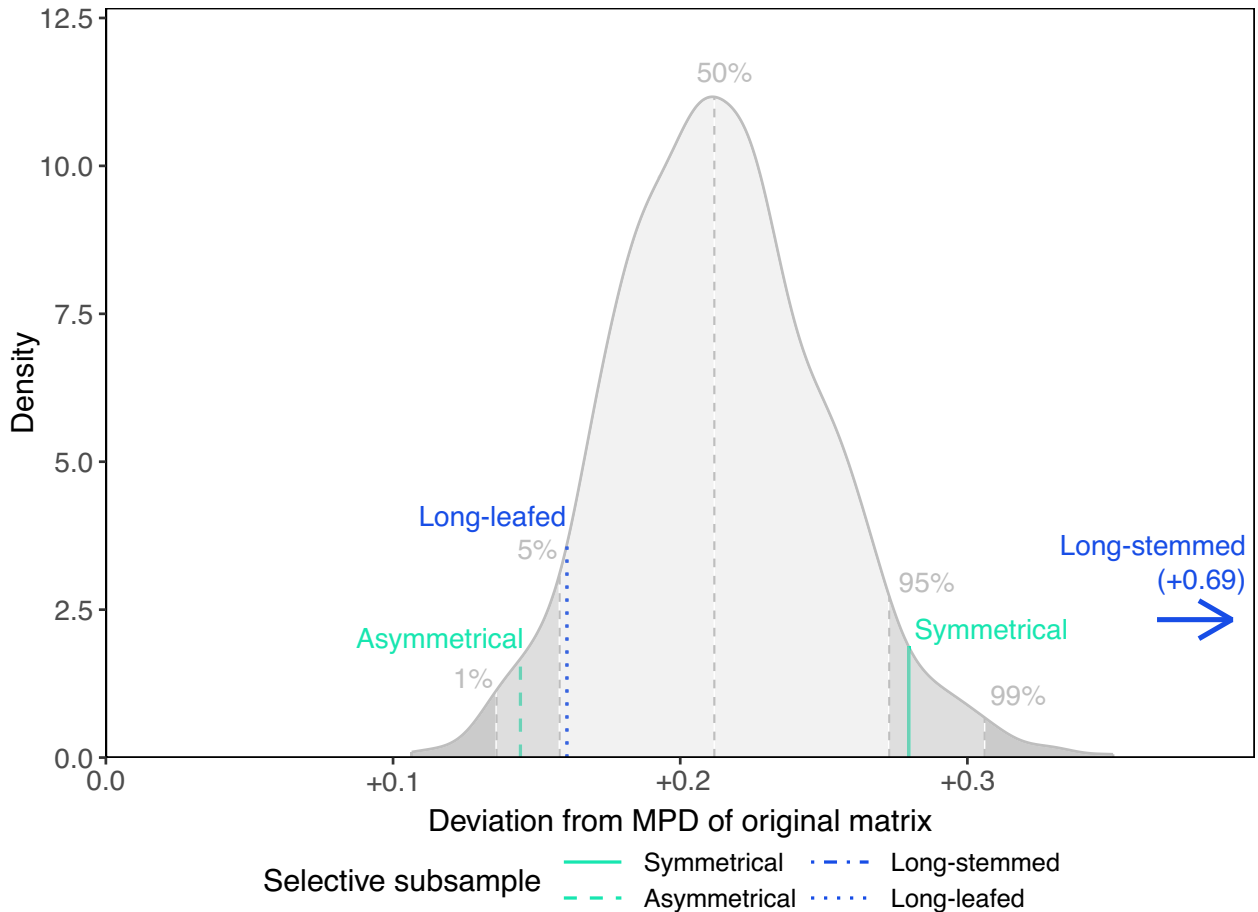


FIG. 10. Deviations in mean pairwise distance of 1000 random and four selective subsamples from that of the original coelurosaur dataset. This plot graphically depicts some of the results presented in Table 3 for ease of understanding. The regions of the distribution delimited by the 1%, 5%, 95%, and 99% dashed lines represent the lower 1%, lower 5%, upper 5% and upper 1% quantiles respectively. The 50% line represents the median deviation. As the long-stemmed subsample deviated substantially more from the original matrix than the other subsamples, its deviation is truncated for aesthetic purposes. Disparity index values of subsampled matrices were scaled to account for differences in total branch length values between subsample subtrees. *Abbreviations:* MPD, mean pairwise distance.

the originals, with the exception of the morphological distance between the most peripheral morphotypes; the aspect of morphospace occupation captured by the sum of ranges index. This is to be expected; while subsampling can increase variance, it can only maintain or decrease range.

The relationship recovered between symmetry and sum of ranges varied between analyses. When datasets simulated along contrasting trees were compared, sum of ranges and tree symmetry exhibited a negative relationship. However, when topological contrasts were introduced through subsampling rather than during data simulation, sum of ranges correlated positively with symmetry. Considering the branch length distributions of the underlying trees resolves this apparent contradiction. The original asymmetrical tree had much longer external branches than both its symmetrical comparator

and the 16-tip replicate generated through subsampling. Based on our results, it is to be expected that the sums of ranges of the matrices simulated along the asymmetrical tree will exceed those of their symmetrical counterparts; an expectation that was borne out when the data were compared. As the subsampling analyses could not replicate these long external branches, the resultant subsamples did not contain any highly divergent peripheral morphotypes, hence a positive relationship between symmetry and sum of ranges was recovered. This relationship is concordant with that displayed by the other indices tested that include some sort of size component in their characterization of morphospace. Taken together, these results demonstrate that tree symmetry cannot be used to predict the divergence of peripheral phenotypes unless differences in branch length distribution are first controlled for.

TABLE 3. Deviation in disparity of selective subsamples emphasizing symmetry and stemminess from the original matrices, compared to deviations of 1000 random subsamples (p-values).

	Mean pairwise distance						Sum of variances						Mean distance from centroid						Sum of ranges					
	Avetheropod		Coelurosaur		Mammal		Crinoid		Avetheropod		Coelurosaur		Mammal		Crinoid		Avetheropod		Coelurosaur		Mammal		Crinoid	
	(higher)	(lower)	(higher)	(lower)	(higher)	(lower)	(higher)	(lower)	(higher)	(lower)	(higher)	(lower)	(higher)	(lower)	(higher)	(lower)	(higher)	(lower)	(higher)	(lower)	(higher)	(lower)	(higher)	(lower)
Symmetrical	0.038*	0.041*	0.041*	0.041*	0.155	0.338	0.028*	0.028*	0.001**	0.001**	0.008*	0.008*	0.085	0.048*	0.048*	0.227	0.011*	0.011*	0.098	0.098	0.349	0.349	0.349	0.349
Asymmetrical	0.096	0.024*	0.076	0.402	0.076	0.402	0.009**	0.009**	0.408	0.009**	0.332	0.341	0.215	0.399	0.399	0.337	0.43	0.43	0.221	0.221	0.486	0.486	0.486	0.486
Long-stemmed	<0.001**	<0.001**	<0.001**	<0.001**	<0.001**	<0.001**	<0.001**	<0.001**	0.083	0.083	<0.001**	<0.001**	0.385	<0.001**	<0.001**	<0.001**	<0.001**	<0.001**	0.372	0.372	<0.001**	<0.001**	<0.001**	<0.001**
Long-leaved	0.001**	0.056	<0.001**	<0.001**	<0.001**	<0.001**	0.046*	0.338	0.338	0.049*	0.018*	0.391	0.124	0.002**	0.002**	0.008**	0.335	0.335	0.112	0.112	<0.001**	<0.001**	<0.001**	<0.001**

*0.05 significance threshold met or exceeded; **0.01 significance threshold met or exceeded.

Selective subsamples labelled 'higher' deviated more positively than the majority of random subsamples, while those labelled 'lower' deviated more negatively.

When both trees were ultrametric, the symmetry of the tree predicted the contribution of nodes to perceptions of evolving morphological disparity. Along the symmetrical tree, the inclusion of nodes in the analysis returned a gradual dispersal across morphospace through time. By the penultimate time partition, the size and density of the area in morphospace occupied by the nodes almost matched that of the tips. The nodes of the asymmetrical tree presented no such gradient; their morphospace occupation remained compact until the inclusion of the tips. The gradient in morphospace occupation through time introduced through the inclusion of bisected lineages bisected by the upper limits of each partition suggests that pairwise phylogenetic distance may be predictive of changes in dispersion through time. However, this possibility was not directly assessed and is beyond the scope of this study.

Branch length distribution. Tree stemminess predicted dispersion within morphospace. This effect could not be replicated through subsampling; while the polarity of the relationship could be reproduced by drawing subsets with relatively higher or lower stemminess, the magnitude of the difference between the original long-stemmed and long-leaved matrices could not. As subsampling cannot increase the overall internal branch length of a tree, this is to be expected. Additionally, the similarity in morphospace occupation between the original long-leaved matrices and the contrasting subsamples derived from them, despite the substantial differences in overall external branch length, further demonstrates how dispersion within morphospace is predominantly defined by character evolution captured by stem branches.

Branch length distribution did not predict the contributions of tips, nodes or bisected lineages to perceptions of evolving morphospace occupation through time. The only differences recovered were rooted in variations in taxonomic diversity; a consequence of tree length inequality and lineage duration heterogeneity across each comparison. While rarefaction can normalize diversity across time series, diverse partitions will still generally yield more heterogeneous subsets than depauperate ones. Consequently, the differences recovered in the disparity-through-time analyses conducted using the paired contrasts in ultrametricity and stemminess reflect the diversity of tips, bisected lineages, and nodes in each partition, rather than genuine evolutionary phenomena.

Empirical data analysis

When subtrees and corresponding matrix subsets were selectively drawn from each dataset to emphasize and minimize symmetry, only the symmetrical subsample of the coelurosaur dataset exhibited a substantial increase in

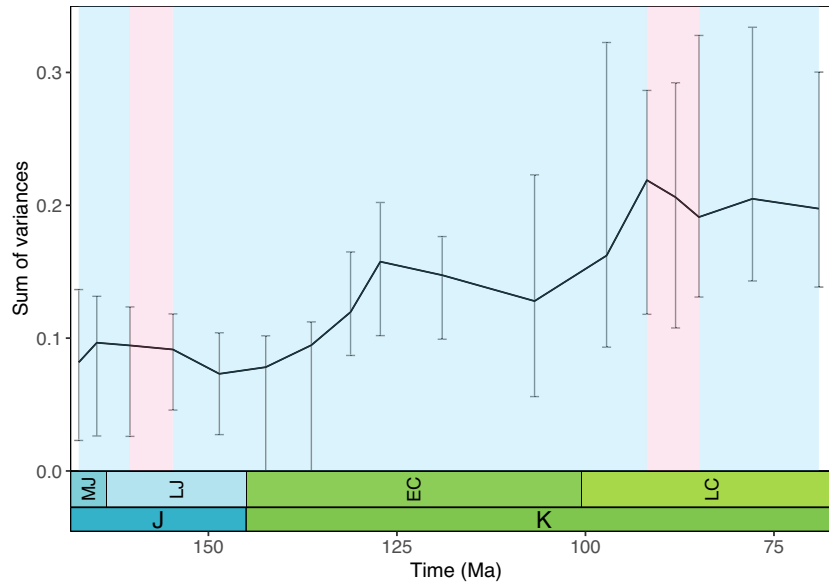


FIG. 11. Sum-of-variances-through-time of coelurosaur dataset partitioned by stratigraphic age. This plot graphically depicts some of the results presented in Table 4 for ease of understanding. Error bars present the 90% confidence interval of 100 bootstraps; the bold line connects the median sum of variance values of each partition. Shading of intervals displays whether the changes in median sum of variances between samples follow predictions informed by variations in partition subtree symmetry. Intervals presenting positive covariance between the sum of variances and tree symmetry are shaded blue; intervals presenting negative covariance are shaded red. Tables 4 and 5 present the results of these analyses for all combinations of dataset, index and tree parameter.

TABLE 4. Percentage of stratigraphic intervals where disparity indices positively covaried with subtree symmetry.

	Avetheropod (%)	Coelurosaur (%)	Mammal (%)	Crinoid (%)
Mean pairwise distance	61.54	81.25	71.43	66.67
Sum of variances	54.55	81.25	62.50	66.67
Mean distance from centroid	63.64	81.25	62.50	66.67
Sum of ranges	54.55	81.25	62.50	73.33

TABLE 5. Percentage of stratigraphic intervals where disparity indices positively covaried with subtree stemminess.

	Avetheropod (%)	Coelurosaur (%)	Mammal (%)	Crinoid (%)
Mean pairwise distance	30.77	50.00	71.43	43.75
Sum of variances	36.36	50.00	50.00	43.75
Mean distance from centroid	36.36	50.00	50.00	43.75
Sum of ranges	36.36	46.15	50.00	37.50

dispersion solely predicted by its topological balance; variations in stemminess were predictive for all other results with $p \leq 0.05$ (see below for further details). Stemminess correlated with increasing morphospace dispersion, but only after differences in total branch length between subsampled trees had been accounted for. In the analyses that applied this correction, the majority of subsets that exhibited a substantial deviation in morphospace occupation from the random subsamples also displayed a marked increase or decrease in stemminess (all significant deviations except for the mean pairwise distance of the symmetrical coelurosaur subsample, see Table 3). This included all subsamples taken with the intention of emphasizing other aspects of tree shape, with the exception of the aforementioned symmetrical coelurosaur

subset. In our empirical disparity-through-time analyses, symmetry consistently and positively covaried with morphospace dispersion. In contrast, stemminess exhibited negative or no covariation with morphospace occupation, depending on the dataset analysed.

Consensus of simulated and empirical data analyses in relation to sampling

Taken together, the analyses based on simulated and empirical data demonstrate that tree symmetry and stemminess can predict the impact of sampling on morphospace occupation. It therefore follows that sampling practices that introduce tendencies in datasets towards

specific tree shapes have the potential to bias morphospaces towards specific distributions. If sampling introduces marked differences in tree symmetry between clades, our results suggest that artefactual differences in morphospace occupation will be introduced. Specifically, clades with more balanced topologies may appear more dispersed within morphospace than those with greater imbalance. For these artefacts to manifest, the contrast in tree symmetry between clades must be pronounced; subtle differences do not introduce detectable signals. Comparisons of the results of our simulated and empirical analyses support this notion; the marked difference in symmetry between the generating trees was evident when the simulated data were compared but was not replicated when the empirical datasets were selectively subsampled. As all four empirical trees were variably imbalanced to begin with, the subtrees derived through selective subsampling exhibited less divergence in terms of topological balance than the contrasting trees used in the simulations, which were fully symmetrical and asymmetrical respectively. As such, it is unsurprising that the simulations recovered a strong predictive relationship between tree symmetry and morphospace dispersion while the empirical analyses did not.

Tree symmetry generally increases as the number of tips decreases (Mooers & Heard 1997), therefore, it is to be expected that smaller samples will have a higher likelihood of presenting more balanced topologies. This expectation was borne out in our empirical disparity-through-time analyses; positive covariation between symmetry and morphospace dispersion through time was observed across all four datasets. This has implications for the interpretation of disparity-through-time analyses, as subtle biases in sampling between partitions introduced during dataset assembly may affect perceptions of evolving morphospace occupation. Furthermore, tree symmetry has a predictive relationship with the contribution of nodes to perceptions of evolving disparity-through-time; along balanced trees, nodes present a gradual increase in dispersion across morphospace, whereas along imbalanced topologies, the distribution of nodes remains compact and static.

Sampling approaches that introduce differences in stemminess between clades can create perceived differences in otherwise comparable patterns of morphospace occupation. As with tree symmetry, increases in stemminess predict greater dispersion within morphospace while decreases predict compaction. However, this compaction might not be evident when the size of the area of explored morphospace is characterized, as highly divergent peripheral morphotypes can drastically expand the envelope of occupation. Long external branches predict the presence of these peripheral morphotypes. Sampling biases reflected in stemminess have a much stronger effect

on morphospaces than those captured by the symmetry of a tree. Subsampling or partitioning a dataset is unlikely to introduce biases reflected in stemminess as such an approach also introduces differences in total branch length between subsets. The inconsistent covariation between stemminess and morphospace dispersion presented by our empirical disparity-through-time analyses demonstrate this. Accounting for these differences, as we did in our selective subsampling of the empirical datasets, reveals a strong predictive relationship between stemminess and morphospace dispersion, but one that is unlikely to manifest in standard analyses of empirical data.

The recovery of comparable trends in disparity-through-time between datasets simulated along contrasting ultrametric and non-ultrametric trees suggests that the application of ancestral state estimation methods to some extant datasets may provide sufficient data to infer historical trends comparable to those seen in palaeontological data. To date, such methods have been used to supplement palaeontological data in analyses of morphological disparity (Brusatte *et al.* 2011), but not as a substitute. Further experimentation with a variety of tree shapes and datasets is required to thoroughly assess the viability of this approach. However, it at least appears plausible that ancestral state estimation could give utility to some neontological datasets in studies of disparity-through-time.

Artefacts of sampling

Tree shape reflects taxon and character sampling. As cladistic datasets are constructed for the purpose of grouping taxa accordingly to their morphological similarity, it is to be expected that the taxa and characters assembled differ from those that would have been included had the intention been to distinguish between them, as is the case in analyses of disparity (Gerber 2019). As such, some sampling bias between clades characterized in a dataset is probably unavoidable. The evaluation of underlying tree and subtree shapes can help identify these areas of insufficient sampling and inform mitigatory measures. However, there are a number of practices adopted during the assembly of cladistic datasets that have the potential to introduce stronger, more concerning artefacts in analyses of morphospace occupation.

A common occurrence in phylogenetic discrete character matrices is the different intensities with which the ingroup and outgroups are sampled (e.g. Prentice *et al.* 2011; Romano 2019). Typically, the ingroup will be sampled more exhaustively than the outgroup since it is, by definition, the focus of any phylogenetic analysis; the outgroup is sampled only to establish evolutionary polarity and its membership can be little more than perfunctory.

This results in trees that are more asymmetric than would be expected from complete, uniform, or random sampling of taxonomic diversity, with the asymmetry concentrated near the root. Consequently, even if the overall diversities of the ingroup and outgroup are equal, our results indicate that the outgroup will present an artefactually compacted pattern of morphospace occupation while the ingroup will appear relatively dispersed. However, as outgroups are reliably excluded from cladistic datasets employed in analyses of morphological disparity, so too are the biases associated with their inclusion.

The inclusion or exclusion of fossil taxa can also have implications for perceptions of morphospace occupation. Since trees of extant species are typically more balanced than trees of extinct species (Harcourt-Brown *et al.* 2001), extant members of clades may appear more morphologically diverse than their extinct relatives by virtue of their geological age alone. Taxonomic rank heterogeneity, commonplace in cladistic matrices (e.g. Prentice *et al.* 2011; Brusatte *et al.* 2014; Bapst *et al.* 2016), is another product of taxon sampling that is reflected in the branch length distribution of a tree. The collapse of diverse lineages into monotypic terminals simultaneously reduces the stemminess of the overall tree while also introducing long external branches. Consequently, this practice is expected to bias clades towards compact distributions within morphospace orbited by highly divergent peripheral morphotypes.

As with taxon sampling, tree shape also reflects the character composition of a dataset, particularly in terms of branch length distribution. When branch length quantifies the amount of evolutionary change taking place, different edges capture different types of character evolution. Coarsely, internal stem branches characterize synapomorphic character evolution while external leaf branches express the evolution of autapomorphies. Following this logic, our results demonstrate that stem branch length predicts the dispersal of taxa within morphospace while leaf branch length predicts the margin by which peripheral morphotypes will diverge.

Autapomorphies are uninformative in parsimony-based phylogenetic analyses; a fact that has led to the perception that their undersampling is to be expected in cladistic datasets (Gerber 2019). Such a deficiency would be reflected in the truncation of what would otherwise be relatively long external branches and would manifest as a morphospace that underestimates the divergence of peripheral morphotypes. However, it is now apparent that the sampling of autapomorphies in cladistic datasets may be more representative than previously thought (Gerber 2019). Similarly, homoplastic characters may not be as poorly sampled in cladistic datasets as is often assumed (Wagner 2012; Hughes *et al.* 2013; Gerber 2019). Nevertheless, undersampling remains a potential issue that

should be assessed and addressed. How homoplasy shapes morphospace is poorly understood. As such, our results cannot speak to the impact of adding, omitting or modifying such characters, with the exception of one specific sampling practice: when a homoplastic character is broken down into a series of autapomorphies. As this practice reassigns character evolution from internal to external branches, the former will decrease in length while the latter increases. Such changes in tree shape predict a morphospace that underestimates dispersion and overestimates peripheral morphotype variance.

Beyond decisions made during dataset assembly, it must be recognized that the non-uniformity of the fossil record precludes representative sampling of morphological diversity. Preservation varies among anatomical features (e.g. Kidwell & Flessa 1995; Behrensmeyer *et al.* 2000), depositional environments (e.g. Patzkowsky & Holland 2012), sedimentary basins (e.g. Holland 2016) and through time (e.g. Foote *et al.* 2007). Consequently, it is to be expected that for most clades, entirely faithful characterizations of the evolution of morphological disparity are not possible.

Biological trends

While the contribution of synapomorphies is relatively well understood (Guillermé *et al.* 2020a), the role that autapomorphies play in shaping morphospaces is less settled. Historically considered to be essential for understanding trends in evolving phenotypic variance through time (Gould 1991), recent analyses have suggested the opposite, that autapomorphies have little impact on morphospace occupation (Cisneros & Ruta 2010; Ruta & Wills 2016; Deline *et al.* 2018). Our results demonstrate that their contribution is subtle; autapomorphic characters contribute little to the overall dispersal of taxa within morphospace, which is predominantly characterized by synapomorphic character evolution. However, including autapomorphies does expand the overall size of the area explored by increasing the range of phenotypes present in each dimension post-ordination. Autapomorphies and synapomorphies characterize different types of morphospace; synapomorphic subspace is defined by character combinations that are theoretically achievable by multiple taxa, whereas autapomorphic subspace can only be traversed by a single taxon by virtue of the defining trait being unique. Large amounts of autapomorphic subspace can create the false perception that a morphospace is undersaturated, even when the synapomorphic subspace of a clade is well explored. It is important to distinguish between such a distribution and one borne out of genuine undersaturation of the synapomorphic subspace, as each has different implications for the evolutionary of

a clade. Assessing the tree shape of a dataset in conjunction with the morphospace it characterizes can help in this regard, as long external branches can serve as indicators for peripheral morphotypes.

Instances of convergence and reversal are important evolutionary phenomena that should be taken into account in analyses of morphological disparity (Foote 1996b; Gerber 2019). However, as the general perception of homoplastic characters is that they are uninformative and potentially misleading in analyses of phylogeny, it is to be expected that these phenomena are inconsistently codified in cladistic datasets (Wagner 2012; Hughes *et al.* 2013). Instead, they may be completely ignored or codified as series of characters differentiated by assumptions of homology. Depending on the prevalence of the homoplastic trait, codification may produce a series of autapomorphies, synapomorphies, or a mix of the two. Our results suggest that alterations to the autapomorphy:synapomorphy ratio of a discrete character dataset will manifest in the morphospace it characterizes. Beyond this simplistic prediction, however, we cannot speak to the impact of homoplastic characters on morphospace as the complex histories of character evolution they imply cannot be inferred from tree shape. Nevertheless, poor understanding of the relationship between homoplasy and morphospace does not justify the absence of characters that codify evolutionary convergence and reversal in datasets intended for disparity analysis.

Index choice

Of the three indices employed to characterize the area of morphospace occupied by a clade, sum of ranges was the only one to collapse under subsampling. This result expands upon the body of evidence highlighting the volatility of range-based indices, such as sum of ranges, and the caution that should be taken when they are employed to characterize morphospace occupation (see Guillerme *et al.* 2020b). While their application can provide useful insights into the phenotypic divergence of peripheral morphotypes, they should be employed in conjunction with other indices that characterize the size of the occupied area of morphospace by other measures.

Mean pairwise distance, sum of variances, and mean distance from centroid covaried across all analyses of the simulated data, but not those exploring the empirical matrices. This is consistent with the notion that there is some overlap in aspects of the morphospace they characterize but that the indices are ultimately distinct in their meaning; mean pairwise distance best characterizes density, while sum of variances and mean distance from centroid measure size but also take the spacing of taxa into account to some degree. As anticipated by Wills (2001), the intertaxon

spacing of empirical samples with high morphological variance was not consistently characterized by their sum of variances in a representative way. These inconsistencies are evident when the mean pairwise distance of the samples is used to characterize the density with which they occupy morphospace, which speaks to the importance of employing multiple indices in analyses of disparity. The presence of this inconsistent relationship between mean pairwise distance, an index of density, and the sum of variances and mean distance from centroid suggests that classification of the latter as indices exclusively characterizing size or density may be misrepresentative, and that it may be pertinent to consider them to be measures of dispersion as originally intended (Foote 1992).

Recommendations for disparity analysis

Disparity analyses are powerful tools for gleaning insights into the structure and evolution of morphological diversity. However, as morphological disparity is comparative, wanton application of these methods to available datasets does little to illuminate the evolutionary history of clades without context. Hypothesis-driven approaches remain the most powerful application of disparity methods, for within well-designed, interrogative frameworks, such analyses yield results with genuine meaning, rather than inductive characterizations of morphological diversity. In keeping with this targeted approach, datasets analysed in disparity frameworks should be bespoke, assembled specifically for the question at hand. However, the assembly of purpose-built disparity datasets may not always be practical due to time constraints, source data availability, and other such limiting factors. In such cases, pre-existing cladistic datasets can constitute a valuable source of information for analyses of morphospace occupation. However, the taxon and character-sampling strategies that inform their construction, recognizable in the tree shape of the data, undermine their uncritical repurposing and exacerbate the problem of fossil record non-uniformity. Our results provide corroborative support for existing proposals of best practice in preparing cladistic datasets for disparity analysis (Hughes *et al.* 2013; Lloyd 2016; Gerber 2019). Prior to analysis, phylogenetic outgroups should be removed from the dataset. Taxonomic coverage should be evaluated to achieve proportional sampling of morphological diversity, either through the addition or removal of taxa. Depending on the question being asked of the data, this evaluation should consider taxonomic coverage as a whole and through time. Taxonomic rank heterogeneity should be addressed either by collapsing tips into higher-rank taxa or by replacing higher taxa with multiple lower-rank subtaxa. Character composition should also be assessed, particularly in terms of potential biases against the inclusion of homoplastic and

autapomorphic characters in cladistic matrices (Gerber 2019). Rather, disparity matrices should seek to proportionately sample characters codifying all types of morphological evolution.

Tree shape can serve as a good indicator of whether taxon sampling will impact on perceptions of morphospace occupation, as characterized by a dataset. If there is reason to suspect such artefacts have been introduced, we recommend that subsampling approaches are adopted to generate distributions of disparity suitable for comparison. While random subsampling may be sufficient to account for biases introduced by peripheral morphotypes and other products of sampling (Guillerme & Cooper 2018), it is important to ensure that the subsamples generated capture a spectrum of tree shapes. If not, it may be pertinent to manually generate subsamples so the interaction between tree shape and taxon content can be directly addressed. Subsampling of the character set may also prove insightful, particularly in disentangling the contributions of different character types to perceptions of morphospace occupation (e.g. Deline *et al.* 2018). Tree shape can be used to infer whether subsampling by character type (i.e. autapomorphies vs synapomorphies) will return meaningful differences in morphospace occupation. Across all analyses, multiple indices of disparity should be employed so that the nuances of the morphospace occupation of each clade can be effectively characterized (Guillerme *et al.* 2020b).

CONCLUSION

Sampling practices employed in the construction of cladistic datasets can introduce artefacts that manifest in analyses of morphological disparity. As tree shape reflects taxon sampling and character composition, it can be used to predict the impact of such artefacts. Specifically, topological symmetry and the apportioning of branch length between internal and external branches can predict patterns of morphospace occupation. Symmetrical trees with relatively long internal branches predict greater dispersion within morphospace, while asymmetrical trees with relatively short internal branches are predictive of more compact distributions. Long external branches predict greater divergence by peripheral morphotypes from the broader clade. Sampling practices that are reflected in the branch length distribution of a tree introduce much stronger biases into analyses of morphospace than those captured by symmetry. However, artefacts reflected in branch length distribution are much less likely to be unintentionally introduced through subsampling than those predicted by symmetry. In disparity-through-time analyses, tree symmetry predicts the contributions of ancestral phenotypes to perceptions of evolving disparity, while branch length distribution provides no such information. Furthermore,

sampling practices that are reflected in branch length distribution appear to have little effect on perceptions of disparity-through-time. This speaks to unexplored potential of ancestral state estimation in giving utility to extant taxon datasets in analyses of historical trends. Collectively, these findings highlight the need for careful consideration of the sampling practices adopted during the construction of cladistic datasets prior to their recycling in analyses of disparity. The artefacts these practices introduce should be assessed and addressed through dataset augmentation and modification. Going one step further, datasets purpose-built for analyses of disparity circumvent all of these issues. When morphological diversity is faithfully represented, differences in tree shape between clades can speak to genuine differences in morphospace occupation. Such representative sampling is essential in analyses of disparity, lest artefacts of sampling be mistaken for genuine evolutionary phenomena.

Acknowledgements. We would like to thank our colleagues in the Bristol Palaeobiology Research Group for discussion. TJS was funded by a Natural Environment Research Council (NERC) PhD Studentship within the GW4+ Doctoral Training Programme. MNP was funded by an 1851 Research Fellowship from the Royal Commission for the Exhibition of 1851. JEO'R, DP and PCJD were funded by NERC grant (NE/P013678/1), part of the Biosphere Evolution, Transitions and Resilience (BETR) programme, which is co-funded by the Natural Science Foundation of China (NSFC); PCJD is also funded by Biotechnology and Biological Sciences Research Council (BB/T012773/1; BB/N000919/1). We would also like to thank Thomas Guillerme, Sylvain Gerber, and a third, anonymous referee for their insightful comments on earlier versions of the manuscript. The final version of the manuscript was much improved by their constructive input.

DATA ARCHIVING STATEMENT

Empirical datasets and R code are available in the Dryad Digital Repository: <https://doi.org/10.5061/dryad.dncjxkz>

Editor. Marcello Ruta

SUPPORTING INFORMATION

Additional Supporting Information can be found in the online version of this article (<https://doi.org/10.1111/pala.12569>):

Appendix S1. Supplementary figures and table.

REFERENCES

ANDERSON, P. S. L. and FRIEDMAN, M. 2012. Using cladistic characters to predict functional variety: experiments

- using early gnathostomes. *Journal of Vertebrate Paleontology*, **32**, 1254–1270.
- BAPST, D. W., WRIGHT, A. M., MATZKE, N. J. and LLOYD, G. T. 2016. Topology, divergence dates, and macroevolutionary inferences vary between different tip-dating approaches applied to fossil theropods (Dinosauria). *Biology Letters*, **12**, 20160237.
- BARIDO-SOTTANI, J., SAUPE, E. E., SMILEY, T. M., SOUL, L. C., WRIGHT, A. M. and WARNOCK, R. C. M. 2020. Seven rules for simulations in paleobiology. *Paleobiology*, **46**, 435–444.
- BECK, R. M. and LEE, M. S. 2014. Ancient dates or accelerated rates? Morphological clocks and the antiquity of placental mammals. *Proceedings of the Royal Society B*, **281**, 20141278.
- BEHRENSMEYER, A. K., KIDWELL, S. M. and GASTALDO, R. A. 2000. Taphonomy and paleobiology. *Paleobiology*, **26**, 103–147.
- BORTOLUSSI, N., DURAND, E., BLUM, M. and FRANCOIS, O. 2006. apTreeshape: statistical analysis of phylogenetic tree shape. *Bioinformatics*, **22**, 363–364.
- BRUSATTE, S. L., LLOYD, G. T., WANG, S. C. and NORELL, M. A. 2014. Gradual assembly of avian body plan culminated in rapid rates of evolution across the dinosaur-bird transition. *Current Biology*, **24**, 2386–2392.
- MONTANARI, S., YI, H. Y. and NORELL, M. A. 2011. Phylogenetic corrections for morphological disparity analysis: new methodology and case studies. *Paleobiology*, **37**, 1–22.
- CAILLIEZ, F. 1983. The analytical solution of the additive constant problem. *Psychometrika*, **48**, 305–308.
- CIAMPAGLIO, C. N., KEMP, M. and MCSHEA, D. W. 2001. Detecting changes in morphospace occupation patterns in the fossil record: characterization and analysis of measures of disparity. *Paleobiology*, **27**, 695–715.
- CISNEROS, J. C. and RUTA, M. 2010. Morphological diversity and biogeography of procolophonids (Amniota: Parareptilia). *Journal of Systematic Palaeontology*, **8**, 607–625.
- COLLESS, D. H. 1982. Review of phylogenetics: the theory and practice of phylogenetic systematics II. *Systematic Zoology*, **31**, 100–104.
- DELINE, B., GREENWOOD, J. M., CLARK, J. W., PUTTICK, M. N., PETERSON, K. J. and DONOGHUE, P. C. J. 2018. Evolution of metazoan morphological disparity. *Proceedings of the National Academy of Sciences*, **115**, E8909–E8918.
- ERWIN, D. H. 2007. Disparity: morphological pattern and developmental context. *Palaeontology*, **50**, 57–73.
- FERRON, H. G., GREENWOOD, J. M., DELINE, B., MARTINEZ-PEREZ, C., BOTELLA, H., SANSOM, R. S., RUTA, M. and DONOGHUE, P. C. J. 2020. Categorical versus geometric morphometric approaches to characterizing the evolution of morphological disparity in Osteostraci (Vertebrata, stem Gnathostomata). *Palaeontology*, **63**, 717–732.
- FIALA, K. I. and SOKAL, R. R. 1985. Factors determining the accuracy of cladogram estimation – evaluation using computer-simulation. *Evolution*, **39**, 609–622.
- FOOTE, M. 1991. Morphological and taxonomic diversity in clade's history: the blastoid record and stochastic simulations. *Contributions from the Museum of Paleontology, The University of Michigan*, **28**, 101–140.
- 1992. Rarefaction analysis of morphological and taxonomic diversity. *Paleobiology*, **18**, 1–16.
- 1996a. Models of morphological diversification. 62–86. In JABLONSKI, D., ERWIN, D. H. and LIPPS, J. H. (eds). *Evolutionary palaeobiology*. University of Chicago Press.
- 1996b. Perspective: evolutionary patterns in the fossil record. *Evolution*, **50**, 1–11.
- 1997. The evolution of morphological diversity. *Annual Review of Ecology & Systematics*, **28**, 129–152.
- CRAMPTON, J. S., BEU, A. G., MARSHALL, B. A., COOPER, R. A., MAXWELL, P. A. and MATCHAM, I. 2007. Rise and fall of species occupancy in Cenozoic fossil mollusks. *Science*, **318**, 1131–1134.
- GAVRILETS, S. 1999. Dynamics of clade diversification on the morphological hypercube. *Proceedings of the Royal Society B*, **266**, 817–824.
- GERBER, S. 2019. Use and misuse of discrete character data for morphospace and disparity analyses. *Palaeontology*, **62**, 305–319.
- EBLE, G. J. and NEIGE, P. 2011. Developmental aspects of morphological disparity dynamics: a simple analytical exploration. *Paleobiology*, **37**, 237–251.
- GOULD, S. J. 1991. The disparity of the Burgess Shale arthropod fauna and the limits of cladistic analysis: why we must strive to quantify morphospace. *Paleobiology*, **17**, 411–423.
- GOWER, J. C. 1971. General coefficient of similarity and some of its properties. *Biometrics*, **27**, 857–871.
- GREEN, W. A., HUNT, G., WING, S. L. and DIMICHELE, W. A. 2011. Does extinction yield an axe or pruning shears? How interactions between phylogeny and ecology affect patterns of extinction. *Paleobiology*, **37**, 72–91.
- GUILLERME, T. 2018. dispRity: a modular R package for measuring disparity. *Methods in Ecology & Evolution*, **9**, 1755–1763.
- and COOPER, N. 2018. Time for a rethink: time subsampling methods indisparity-through-time analyses. *Palaeontology*, **61**, 481–493.
- BRUSATTE, S. L., DAVIS, K. E., JACKSON, A. L., GERBER, S., GOSWAMI, A., HEALY, K., HOPKINS, M. J., JONES, M. E. H., LLOYD, G. T., O'REILLY, J. E., PATE, A., PUTTICK, M. N., RAYFIELD, E. J., SAUPE, E. E., SHERRATT, E., SLATER, G. J., WEISBECKER, V., THOMAS, G. H. and DONOGHUE, P. C. J. 2020a. Disparities in the analysis of morphological disparity. *Biology Letters*, **16**, 8.
- PUTTICK, M. N., MARCY, A. E. and WEISBECKER, V. 2020b. Shifting spaces: which disparity or dissimilarity measurement best summarize occupancy in multidimensional spaces? *Ecology & Evolution*, **10**, 7261–7275.
- HARCOURT-BROWN, K. G., PEARSON, P. N. and WILKINSON, M. 2001. The imbalance of paleontological trees. *Paleobiology*, **27**, 188–204.
- HARMON, L. J., SCHULTE, J. A., LARSON, A. and LOSOS, J. B. 2003. Tempo and mode of evolutionary radiation in iguanian lizards. *Science*, **301**, 961–964.
- HOLLAND, S. M. 2016. The non-uniformity of fossil preservation. *Philosophical Transactions of the Royal Society B*, **371**, 11.
- HOPKINS, M. J. and GERBER, S. 2017. Morphological disparity. In NUÑO DE LA ROSA, L. and MÜLLER, G.

- (eds). *Evolutionary developmental biology: A reference guide*. Springer.
- HUGHES, M., GERBER, S. and WILLS, M. A. 2013. Clades reach highest morphological disparity early in their evolution. *Proceedings of the National Academy of Sciences*, **110**, 13875–13879.
- KIDWELL, S. M. and FLESSA, K. W. 1995. The quality of the fossil record: populations, species, and communities. *Annual Review of Ecology & Systematics*, **26**, 269–299.
- KORN, D., HOPKINS, M. J. and WALTON, S. A. 2013. Extinction space—a method for the quantification and classification of changes in morphospace across extinction boundaries. *Evolution*, **67**, 2795–2810.
- LLOYD, G. T. 2016. Estimating morphological diversity and tempo with discrete character-taxon matrices: implementation, challenges, progress, and future directions. *Biological Journal of the Linnean Society*, **118**, 131–151.
- MOOERS, A. O. and HEARD, S. B. 1997. Inferring evolutionary process from phylogenetic tree shape. *Quarterly Review of Biology*, **72**, 31–54.
- NIKLAS, K. J. 1986. Computer-simulated plant evolution. *Scientific American*, **254**, 78–86.
- O'REILLY, J. E., PUTTICK, M. N., PARRY, L., TANNER, A. R., TARVER, J. E., FLEMING, J., PISANI, D. and DONOGHUE, P. C. 2016. Bayesian methods outperform parsimony but at the expense of precision in the estimation of phylogeny from discrete morphological data. *Biology Letters*, **12**, 20160081.
- PARADIS, E. and SCHLIEP, K. 2019. ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. *Bioinformatics*, **35**, 526–528.
- PATZKOWSKY, M. E. and HOLLAND, S. M. 2012. *Stratigraphic paleobiology: Understanding the distribution of fossil taxa in time and space*. University of Chicago Press.
- PIE, M. R. and WEITZ, J. S. 2005. A null model of morphospace occupation. *American Naturalist*, **166**, E1–E13.
- PRENTICE, K. C., RUTA, M. and BENTON, M. J. 2011. Evolution of morphological disparity in pterosaurs. *Journal of Systematic Palaeontology*, **9**, 337–353.
- PUTTICK, M. N., GUILLERME, T. and WILLS, M. A. 2020. The complex effects of mass extinctions on morphological disparity. *Evolution*, **74** (10), 2207–2220.
- R CORE TEAM. 2020. R: A language and environment for statistical computing. v3.6.3, R Foundation for Statistical Computing. www.R-project.org
- REVELL, L. J. 2012. phytools: an R package for phylogenetic comparative biology (and other things). *Methods in Ecology & Evolution*, **3**, 217–223.
- RODGERS, J. S. 1993. Response of Colless's tree imbalance to number of terminal taxa. *Systematic Biology*, **42**, 102–105.
- ROMANO, M. 2019. Disparity vs. diversity in Stegosauria (Dinosauria, Ornithischia): cranial and post-cranial sub-dataset provide different signals. *Historical Biology*, **31**, 857–865.
- RUTA, M. and WILLS, M. A. 2016. Comparable disparity in the appendicular skeleton across the fish-tetrapod transition, and the morphological gap between fish and tetrapod postcrania. *Palaeontology*, **59**, 249–267.
- SANDERSON, M. J. and DONOGHUE, M. J. 1989. Patterns of variation in levels of homoplasy. *Evolution*, **43**, 1781–1795.
- SCHAEFFER, J., BENTON, M. J., RAYFIELD, E. J. and STUBBS, T. L. 2020. Morphological disparity in theropod jaws: comparing discrete characters and geometric morphometrics. *Palaeontology*, **63**, 283–299.
- SCHLIEP, K. P. 2011. phangorn: phylogenetic analysis in R. *Bioinformatics*, **27**, 592–593.
- SMITH, T. J., PUTTICK, M. N., O'REILLY, J. E., PISANI, D. and DONOGHUE, P. C. J. 2021. Data from: Phylogenetic sampling affects evolutionary patterns of morphological disparity. *Dryad Digital Repository*. <https://doi.org/10.5061/dryad.dncjsxkzw>
- SUTHERLAND, J. T. F., MOON, B. C., STUBBS, T. L. and BENTON, M. J. 2019. Does exceptional preservation distort our view of disparity in the fossil record? *Proceedings of the Royal Society B*, **286**, 10.
- WAGNER, P. J. 2012. Modelling rate distributions using character compatibility: implications for morphological evolution among fossil invertebrates. *Biology Letters*, **8**, 143–146.
- WICKHAM, H. 2016. *ggplot2: Elegant graphics for data analysis*. Springer.
- WILLS, M. A. 2001. Morphological disparity: a primer. 55–144. In ADRAIN, J. M., EDGEcombe, G. D. and LIEBERMAN, D. S. (eds). *Fossils, phylogeny, and form: An analytical approach*. Springer.
- WRIGHT, D. F. 2017. Bayesian estimation of fossil phylogenies and the evolution of early to middle Paleozoic crinoids (Echinodermata). *Journal of Paleontology*, **91**, 799–814.