BIOLOGY LETTERS

royalsocietypublishing.org/journal/rsbl

Review

Cite this article: Guillerme T *et al*. 2020 Disparities in the analysis of morphological disparity. *Biol. Lett.* **16**: 20200199.
http://dx.doi.org/10.1098/rsbl.2020.0199

Received: 30 March 2020
Accepted: 5 June 2020

Subject Areas:
evolution, palaeontology, ecology

Keywords:
multidimensionality, palaeobiology, ecology, morphology, disparity, variance/variation

Author for correspondence:
Thomas Guillerme
e-mail: guillert@tcd.ie

Palaeontology

Disparities in the analysis of morphological disparity

Thomas Guillerme[1,18,†], Natalie Cooper[2,†], Stephen L. Brusatte[3], Katie E. Davis[4], Andrew L. Jackson[5], Sylvain Gerber[6], Anjali Goswami[2], Kevin Healy[7], Melanie J. Hopkins[8], Marc E. H. Jones[9], Graeme T. Lloyd[10], Joseph E. O'Reilly[11,12], Abi Pate[12], Mark N. Puttick[12,13], Emily J. Rayfield[12], Erin E. Saupe[14], Emma Sherratt[15], Graham J. Slater[16], Vera Weisbecker[1,17], Gavin H. Thomas[18] and Philip C. J. Donoghue[12]

[1]School of Biological Sciences, University of Queensland, St Lucia, QLD4072, Australia
[2]Department of Life Sciences, Natural History Museum London, Cromwell Road, London SW7 5BD, UK
[3]School of GeoSciences, University of Edinburgh, Grant Institute, Edinburgh EH9 3FE, UK
[4]Leverhulme Centre for Anthropocene Biodiversity, Department of Biology, University of York, York YO10 5DD, UK
[5]Department of Zoology, School of Natural Sciences, Trinity College Dublin, Dublin 2, Ireland
[6]Institut de Systématique, Évolution, Biodiversité (ISYEB), Muséum national d'Histoire naturelle, CNRS, Sorbonne Université, EPHE, Université des Antilles, 57 rue Cuvier CP39, 75005 Paris, France
[7]Ryan Institute, School of Natural Sciences, National University of Ireland, Galway, H91YD6H, Ireland
[8]Division of Paleontology (Invertebrates), American Museum of Natural History, New York, NY 10024, USA
[9]Department of Cell and Developmental Biology, University College London, London WC1E 6BT, UK
[10]School of Earth and Environment, University of Leeds, Leeds LS2 9JT, UK
[11]MRC Institute of Genetic and Molecular Medicine, University of Edinburgh, Edinburgh EH4 2XU, UK
[12]School of Earth Sciences, University of Bristol, Life Sciences Building, Tyndall Avenue, Bristol BS8 1TQ, UK
[13]Milner Centre for Evolution, University of Bath, Bath BA2 7AY UK
[14]Department of Earth Sciences, University of Oxford, S Parks Road, Oxford OX1 3AN, UK
[15]School of Biological Sciences, The University of Adelaide, Adelaide, South Australia 5005, Australia
[16]Department of the Geophysical Sciences, University of Chicago, Chicago, IL 60637, USA
[17]College of Science and Engineering, Flinders University, Adelaide, South Australia 5042, Australia
[18]Department of Animal and Plant Sciences, The University of Sheffield, Sheffield S10 2TN, UK

TG, 0000-0003-4325-1275; NC, 0000-0003-4919-8655; SLB, 0000-0001-7525-7319; KED, 0000-0002-9235-7853; ALJ, 0000-0001-7334-0434; SG, 0000-0003-3155-5722; AG, 0000-0001-9465-810X; KH, 0000-0002-3548-6253; MJH, 0000-0002-3580-2172; MEHJ, 0000-0002-0146-9623; GTL, 0000-0001-6887-3981; JEO, 0000-0001-9775-253X; MNP, 0000-0002-1011-3442; EJR, 0000-0002-2618-750X; EES, 0000-0002-0370-9897; ES, 0000-0003-2164-7877; GJS, 0000-0002-6306-0911; VW, 0000-0003-2370-4046; GHT, 0000-0002-1982-6051; PCJD, 0000-0003-3116-7463

Analyses of morphological disparity have been used to characterize and investigate the evolution of variation in the anatomy, function and ecology of organisms since the 1980s. While a diversity of methods have been employed, it is unclear whether they provide equivalent insights. Here, we review the most commonly used approaches for characterizing and analysing morphological disparity, all of which have associated limitations that, if ignored, can lead to misinterpretation. We propose best practice guidelines for disparity analyses, while noting that there can be no 'one-size-fits-all' approach. The available tools should always be used in the context of a specific biological question that will determine data and method selection at every stage of the analysis.

[†]These authors contributed equally to the study.

THE ROYAL SOCIETY PUBLISHING

1. Introduction

Clades of organisms are characterized by variation in both numbers of species and range of phenotypes through time. At the extremes, clades may be exceptionally rich in species and phenotypic diversity (hereafter *disparity*) (e.g. cichlids or molluscs), species-rich but disparity-poor (e.g. rodents or nematodes), species-poor but rich in disparity (e.g. afrotherian mammals) or depauperate in both species diversity and disparity (e.g. lungfish). These phenomena suggest that taxonomic diversity and phenotypic disparity are not inextricably linked, raising important questions, such as: how does disparity evolve? Are some morphologies more common than others? Is anatomical evolution unbounded or are some anatomies impossible to achieve? What role does ecology play in structuring disparity? Analyses of species diversity have a venerable history, but those of disparity are comparatively more recent. Originally defined as 'multidimensional morphological dissimilarity at a macroevolutionary scale' [1,2], the concept of disparity emerged from attempts by palaeobiologists to characterize the evolutionary origin of animal bodyplans and from attempts by comparative developmental biologists to provide causal explanations for their emergence. However, disparity analyses have since expanded into comparative biology as a means of capturing how intrinsic and extrinsic causal agents affect morphological evolution. Typically, methods to capture disparity are based on multidimensional spaces where each dimension represents an aspect of morphological variation (a trait) and biological observations (e.g. taxa) can be placed in this space based on their trait values. Such multidimensional spaces (or morphospaces—defined broadly hereafter as a mathematical space relating morphological configurations generally based on some measure of similarity [3]) can then be used to tackle a diverse array of questions that can be grouped into four main (non-mutually exclusive) classes.

(a) Descriptive disparity

Pioneering studies of disparity characterized the shapes of organisms and how they differed among groups [4,5]. These studies described multidimensional patterns in morphological trait diversity by addressing pertinent questions: why are some morphological trait combinations more common than others, and what are the biological (or mathematical) properties of the resulting morphospace? [4,6,7]. More recently, this approach has been used to understand the relationship between developmental processes and morphology in the field of evolutionary development (evo-devo). For example, patterns of disparity have been used successfully to compare modules of evolution in various groups [8,9], allowing researchers to link variation in shape to a group's evolutionary or developmental constraints [10].

(b) Disparity through time

This approach investigates how the morphologies of organisms have changed over time, by focusing on the disparity of taxa in particular time intervals or slices. This approach has been used widely in palaeobiology to answer a range of macroevolutionary questions, such as: how does disparity accumulate over the history of a clade [11–13], or how does disparity change up to and across mass extinction events [14]?

(c) Disparity and taxonomic diversity

Morphological disparity provides another perspective on biodiversity; high morphological disparity represents a high diversity of morphologies (i.e. shapes or body plans) and is, presumably, associated with high levels of ecological and functional diversity (but see [15]). This makes disparity an informative complement to diversity measures based on species richness alone. Indeed, most studies that have investigated disparity and taxonomic diversity support an effective decoupling of the two (e.g. [16,17]). The approach has been used to investigate whether some groups are more successful than others in their exploration of new evolutionary strategies [18].

(d) Disparity as a proxy for ecology

The disparity of a group can be used as a proxy for either the functional role it plays within an ecosystem or its ecological niche. This approach assumes that groups with high disparity are also likely to be functionally and ecologically diverse and that groups found in similar regions of shape space will have similar functional and ecological roles [14,18]. The links between form and function, however, are not always clear. Traits can be linked to multiple functions and multiple functions can be linked to a single trait [19]. This approach has been used to investigate hypotheses of competitive replacement [20] and changes in ecosystem function during and after mass extinctions [14]. It is one of the primary ways to investigate ecosystem functioning in palaeobiology when the study species (and their functional characteristics) are extinct [19].

Fundamental insights into evolutionary biology have been elicited from these four types of disparity analysis. One of the most important insights is the discovery that morphological disparity is often greatest early in the evolutionary history of clades [21–23], indicating that capacity for evolutionary innovation wanes as clades age, which some have argued reflects the evolutionary assembly of gene regulatory networks that constrain later fundamental change [22,23]. However, this example also highlights one of the greatest challenges confronting researchers who are attempting, increasingly, to obtain general insights from multiple independent studies: can the insights gained from studies using a diversity of methods, approaches and data types be considered equivalent?

In attempting to answer this question, we review current methods and highlight their limitations, as part of a more general attempt to propose best practice guidelines for studies of disparity. We first discuss the appropriate data required for characterising disparity, then review various challenging aspects of these approaches. Throughout, it is important to remember that these tools should always be used in the context of a specific scientific question, as this will drive data and methodological choices at every stage of the process.

2. Data and disparity

Disparity analyses are based on traits, but traits can be characterized in a number of ways: (i) discrete morphological characters, e.g. coding the absence or presence of features or a discrete characteristic of a trait (e.g. [24,25]); (ii) continuous measurements of features (e.g. lengths in [14]); or (iii) more mathematical descriptors from geometric morphometric landmark data (e.g. Procrustes coordinates) (e.g. [26]), Fourier

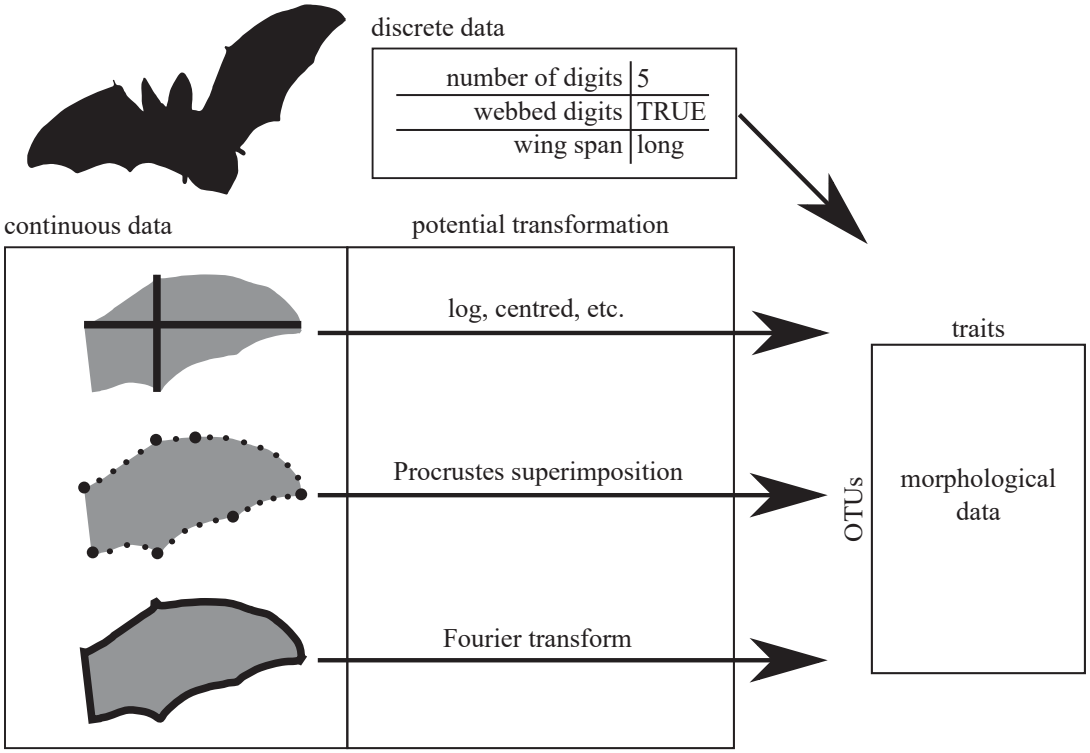

Figure 1. Major routes to obtain morphological data for disparity analyses. Data can be collected as discrete trait observations (e.g. presence–absence data) or as continuous data. Continuous data can be collected by various methods including linear measurements and landmark coordinates or contours (curves). These measurements can then be mathematically transformed (logarithmic transformations, scaling, Procrustes superimposition, elliptic Fourier transforms, etc.). Regardless of the method, data collection produces a trait matrix where the observed traits constitute columns and the studied elements (generally taxa or OTUs) the rows.

coefficients (e.g. [24,27]) or model-based descriptors (e.g. [6,28]; figure 1). None of these approaches is superior, but they may be more or less well suited to characterizing the traits compared and to the question being asked using those traits [29,30].

For example, if investigating variation of bat wing shapes, both homologous landmarks and continuous measurements of bones may be appropriate to capture patterns of wing variation. If the question focuses on comparing wings between bats and birds, however, different measurements might be more appropriate depending on the specific question. That is, if the focus is whether the aerodynamic properties of wings vary within bats or between bats and birds, the traits collected should reflect these aerodynamic properties (e.g. wingspan, aspect ratio, etc.). However, if the focus is on convergence between different bats and birds, it would be preferable to use traits that have facilitated flight in both groups (e.g. digit length, integumentary system, etc.). Where there is any doubt about which traits to analyse, it may be preferable to use several different kinds of data for the same feature to determine whether they capture the same pattern of disparity.

The points above assume that researchers are collecting their own data for disparity analyses, but this is often not the case. Discrete characters are commonly recycled from phylogenetic studies (e.g. [11,31]). This approach may artefactually increase disparity between phylogenetically distinct groups, since phylogenetic characters are often collected to discriminate among groups. This needs to be considered when interpreting results, especially as synapomorphies can lead to apparent shifts or increases in disparity when new clades appear (particularly if the character-state distribution is skewed towards a particular clade). Furthermore, many datasets are limited to subsets of anatomy that are at least implicit samples of overall anatomy, but explicit tests of this assumption have shown that different aspects of morphology can exhibit different patterns of disparity [30]. The influence of trait choice on resulting disparity patterns can be especially challenging where the available data have non-random missing anatomical parts, such as the absence of soft tissue in the fossil record [25].

Ultimately, disparity analyses are characterized by the data they use. Unfortunately, trait data suffer from the same shortcomings as most biological datasets. The data within them can be non-overlapping, hierarchical, inapplicable, ambiguous, polymorphic and/or correlated [32]. There are also issues of missing data, both where a particular character cannot be measured for a given taxon and where a given taxon cannot be sampled at all. Trait data may also be influenced by biological phenomena such as allometry and sexual dimorphism. More practically, data collection is constrained by the time and money available, making collating a 'perfect' dataset impossible. Even when care is taken, subsamples of the universe of possible data may not have the power to uncover the full patterns of disparity. These issues should be considered when collecting data. It is particularly important to collect trait data with the scientific question in mind, or, where there are limits on the data available, to tailor the question being asked to match the data.

3. Disparity analysis methods

Once suitable trait data have been collected, the design of the disparity analysis itself needs to be considered. Study design encompasses several key aspects including (§3a) the difficulty of dealing with multidimensional data; (§3b) the indices used to summarize the relative disparity of groups; (§3c) the methods used for hypothesis testing within the disparity analysis framework; and (§3d) the influence of phylogeny on disparity analyses. We consider these aspects in order below.

(a) To ordinate or not to ordinate? That is the (multidimensional) question

Disparity analyses often use ordination techniques for dimensionality reduction. Ordinations are statistical methods that map observed variables onto a new space of reduced dimension while maintaining the requirement that similar observations are closer together than dissimilar ones (e.g. principal component analysis—PCA; principal coordinates analysis—PCO; non-metric multidimensional scaling—NMDS). They come in many flavours depending on the data and the desired morphospace properties. For example, quantitative (continuous) data can be reduced using PCA, and dissimilarity matrices based on qualitative, quantitative or mixed data types can be reduced using PCO (which is equivalent to metric multidimensional scaling (MDS)) or NMDS (see [33, ch. 9] for a detailed overview of ordination methods and properties). Note that for PCO, the distance metric used can have significant impacts on the resulting morphospace [34]. The choice of distance metric is, therefore, crucial, and should not be overlooked when using PCO.

One of the reasons why ordination techniques are common in disparity analysis is that they make it easier for researchers to comprehend patterns in two or three spatial dimensions at a time, which can be more intuitive than through disparity indices (see §3b). Additionally, after ordinating the data, it is possible to focus on just a subset of axes of the morphospace (i.e. selecting only those axes that describe the majority of the variation in the dataset—e.g. 95%). In the case of geometric morphometric data, some ordination techniques (e.g. PCA) can be particularly useful as they conserve the mathematical properties of the data while efficiently reducing the dimensions [35]. In practice, this facilitates interpretation of only the major axis of a highly dimensional dataset as major gradients of biological variation (e.g. the elongation and flattening of birds' beaks; [36]).

Like most other aspects of disparity analyses, however, reducing dimensionality can be fraught. In the case of ordination, subsampling axes from the ordination can lead to misinterpretation of the results. Although a common technique is to consider the *d*-axes that encompass 95 or 99% of the variance in the dataset (either by manually selecting the *d*-axes that encompass the desired cumulative variance or using methods such as the broken stick model; [33, p. 410], the interpretation of these principal axes can miss some aspects of the structure of the data and lead to misinterpretation of the biological variation mapped on these axes [37,38]. Visual interpretations of multidimensional data can be particularly misleading, not least as multidimensional spaces might not possess the Euclidean properties one often intuitively assumes [7,25].

Interpreting biological variation along the axes is always a *post hoc* procedure and may have little relation to the overall question (for example, if the first few ordination axes represent elongation of the beak in birds, but the question is about wing disparity). Additionally, in some cases, reducing the dimensionality of a dataset can render its interpretation more problematic. For example, when the analysed dissimilarity data are non-Euclidean (e.g. as induced, for instance, with inapplicable characters in discrete character schemes), interpreting the resulting ordinated space can be challenging [39]. This can sometimes be problematic when comparing the position of groups in multidimensional space, as true dissimilarities might not be reliably conveyed (although this can sometimes be improved [40]). Furthermore, *post hoc* interpretations of the gradient of variation on the ordination axes may be biologically meaningless or simply impossible [39]. Although some gradients are easy to detect or interpret (e.g. the elongation and depth of mandibles in fishes on first and second principal components (PC) axes, respectively; [41]), some are not (e.g. [38]). For example, with discrete morphological data, a gradient between the species that have many characters in state 1 and those that have more in state 0 has no biological meaning if these are binary alternate states.

In general, categorical data are a good deal more problematic than continuous data, because the characters themselves are invariably non-equivalent, non-independent and the distribution of the variance is usually more evenly distributed across axes (i.e. contrary to a PCA, the first few axes do not encompass most of the variance in the dataset). Such non-Euclidean spaces often have non-intuitive properties, for example, straight lines viewed in bivariate plots of some dimensions are not actually straight and character coding and missing data can make the pairwise dissimilarity matrix lose its metric properties (i.e. the distance between A and B is not equal to the distance between B and A; [39]). Last, but not least, in many cases, ordination might not be necessary. For example, if an index characterizing disparity can use all of the data, it is not necessary to calculate it on the ordinated dataset (e.g. [31]). For all of these reasons, multidimensional data should not be ordinated automatically, and careful consideration should be given to whether the aim of the study can be achieved without ordination [42,43].

(b) Summarizing disparity using disparity indices

Most disparity datasets are multidimensional and, consequently, a large component of any disparity analysis involves considering how to extract a meaningful (i.e. interpretable) summary of disparity (figure 2). This summary is usually achieved with a disparity measure or index [30]. As with any summary of multidimensional data, disparity indices will reflect only some aspects of the morphological variation, never its whole complexity [46]. It is, therefore, often beneficial to use more than one index to summarize different aspects of variation, guided by the aim of the study.

When considering only one dimension, disparity indices can be used to compare the spread of distributions (e.g. the range, quantiles or variance) or the differences in the central tendencies (i.e. mean, median or mode) of groups in the morphospace. Among these indices, some will have more attractive properties than others, such as sensitivity to outliers. Range and mean are highly sensitive, whereas quantiles, variance and median are less so, making them more or less appropriate for different questions. For example, if the goal is to characterize the extent of morphospace occupied by a group (e.g. does group A occupy as much space as group B?), indices related to the spread of the group in the morphospace are most appropriate (e.g. volume [47]; distance from the centroid [30,48]; variance and range [11]). Furthermore, aspects other than variation (*sensu* disparity) can be of interest: if we wish to describe the 'position' of a group in a morphospace (e.g. does group A occupy the same region of morphospace as group B?), indices related to the distance between the elements within a group and a fixed point in the morphospace are most appropriate [46]. Finally, if we aim to

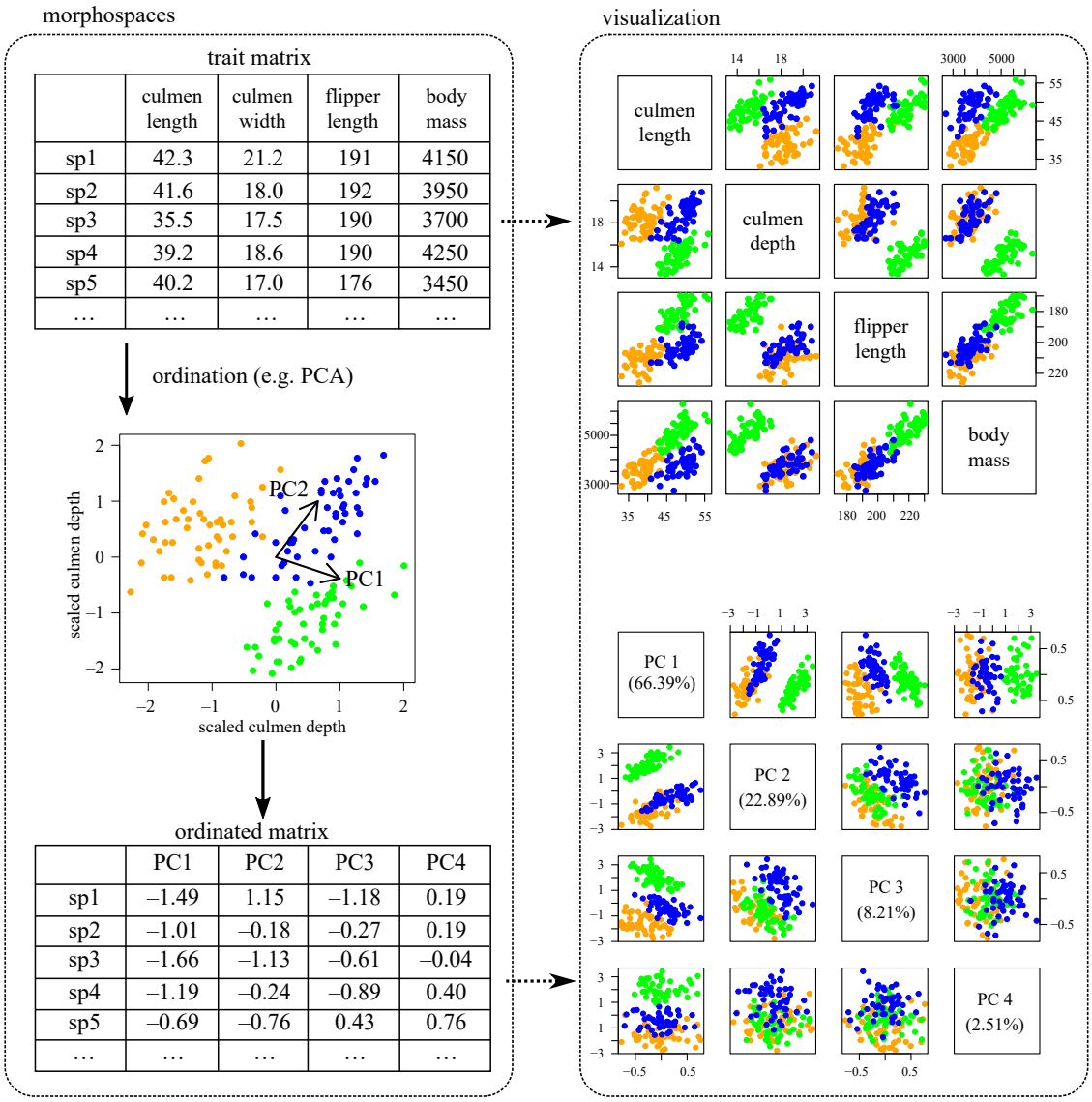

Figure 2. Illustration of the relationships between the different morphospaces and visualization of the same dataset (the 'penguin' dataset of [44,45]). Morphospaces: different mathematical representations of a morphospace. A trait matrix can be an ordinated matrix (e.g. in [20]) or transformed into a distance matrix (e.g. in [31], not represented here). Here, we consider all these matrices as being morphospaces, i.e. objects containing all the combinations of traits and observations (albeit transformed differently). Visualization: different ways to represent the morphospace in 2D. Visualizations can use either trait plots (directly from the trait matrix); or ordination axis plots (directly from the ordinated matrix). Note that in 2D representations, it is good practice to plot both axes on the same scale to avoid visually distorting the importance of one axis.

characterize the density of morphospace occupation (e.g. is group A more closely packed than group B?), indices related to the pairwise distances between elements will be most appropriate (e.g. nearest neighbour distance, pairwise distances, etc. [31]—see §3a).

In addition to considering which properties of disparity these indices capture, it is also important to consider the mathematical properties of the indices and their associated caveats [49,50]. For example, measuring the sum of variance for each dimension of the space before or after ordination via PCA is equivalent. However, this is no longer true of other transformations of the space or when a subset of dimensions or elements are considered, as is often done after PCA [33].

Furthermore, multidimensional spaces have some counter-intuitive properties that should be considered, such as the 'curse of dimensionality' [51]. In spaces with some axes of variance lower than one, product-based indices used as proxies of volumes (e.g. product of ranges, hypervolume, hypercube, etc.) can quickly tend towards zero for spaces with even a modest number of dimensions [51]. Other types of indices are also extremely sensitive to outliers and can be biased easily by sample size, for example, range [49] or convex hull-based [52] indices.

(c) Testing for differences in disparity

No matter which disparity indices have been calculated, the research question must be framed in an appropriate statistical context. The multidimensional statistical toolkit for ecology and evolution has greatly expanded in recent years [53,54], but some of these advances have yet to be implemented in disparity analyses. Instead, hypothesis testing has been mostly confined to a small set of well-established methods. One commonly used test is the non-parametric permutation analysis of variance (PERMANOVA) [55], which tests whether two groups share the same centroid and dispersion based on a distance matrix between observations. The past

decade has also seen a series of developments based on this test (e.g. the linear regression for multidimensional data [56] or the phylogenetic ANOVA [57]; but see [42,54] for more). It is worth noting that most of these tests do not require the morphospace to be ordinated (see §3a). Regardless of the statistical test used, they should only be employed if they are tailored to the question at hand, rather than simply following common practices.

It is also important to consider which data should be subjected to a statistical test. For example, in morphological disparity analysis, especially for palaeobiological questions, data are often bootstrapped. This has two advantages: (i) when the disparity index is unidimensional (e.g. the sum of variances), bootstrapping the data generates a distribution of the index that can be analysed using the vast statistical toolkit available for comparing distributions; (ii) when data are scarce, bootstrapping the data allows users to introduce variance, rendering the test less sensitive to outliers. However, bootstrapped data are pseudoreplicates and thus non-independent and can increase the false positive rate (Type I error) [58]. This, again, highlights the importance of tailoring the statistical test to the data and question at hand.

Finally, it is important to understand the limitations of the dataset for performing statistical analysis. Mainly, disparity analyses should be restrained to groups within the same morphospace and are more difficult between different morphospaces. This can be the case when comparing elements with different numbers of landmarks or different landmark configurations that will result in different morphospaces; comparing disparity indices between these is not trivial.

(d) Disparity and phylogeny

As with all comparative datasets, the data used in disparity analyses are not independent because close relatives will tend to have more similar morphologies than more distant relatives [59]. Thus, for disparity analyses that consider groups with phylogenetic relationships (which is common), the non-independence between observations should be taken into account. It has been noted, however, that some popular phylogenetic correction methods (like phylogenetic PCA) can be inappropriate, especially when using only the first *d*-axes of the ordination, and can lead to incorrect interpretations of the data (such as wrongly supporting 'early burst' type patterns; [60]). Furthermore, any use of phylogenies in disparity analyses must also carefully consider the underlying model of trait evolution. Standard methods assume a model of Brownian motion, i.e. a 'random walk' model where trait variance increases linearly through time with no trend in the direction of trait evolution. In many biological situations, this model is not realistic, and different models of evolution should be considered [61,62]. If an inappropriate model is used then methods such as phylogenetic PCA and ancestral state estimations (see below) may give misleading results, with implications for downstream results of disparity analyses.

One other common way to take phylogeny into account in disparity analyses is using ancestral state estimations in disparity through time analyses to extract disparity estimates for non-sampled taxa and/or nodes of a phylogeny [13,63]. Ancestral state estimation can be performed at two points in the disparity analysis pipeline: either (i) pre-transformation, i.e. the estimation is done before transformation of the data (e.g. ordination, or distance matrix construction) and is simply based on the original data, or (ii) post-transformation, i.e. the estimation is done after transformation of the data by estimating the ancestral states using the transformed matrix (e.g. the ordination scores; [43]).

Pre-transformation ancestral state estimation will change the way the ordination space is defined—i.e. the relationship between the points is not yet estimated—and requires longer computational times. However, once the morphospace is defined, its properties will not change. Post-transformation ancestral state estimation will not change the empirical ordination space and is faster to compute, but it will add elements in the space, whose estimated positions can be problematic for statistical tests and evolutionary inferences down the line [39,43].

All ancestral state estimates are highly dependent on the data and method used (especially on the underlying model of trait evolution) [64]. In general, using ancestral state estimation can help with recovering patterns of change in disparity but should not be used simply to generate extra data points to increase statistical power. In fact, these extra points are not independent and can also have problematic side effects, especially when testing for the influence of mass extinctions on disparity as they artificially and asymmetrically increase taxon sampling.

4. Disparity analyses for the future

Morphological disparity analyses are widely employed in evolutionary palaeobiology, and are based on a diversity of methods and data. There is no 'one-size-fits-all' pipeline for morphological disparity analyses. As with any multidimensional analysis, there are many variables that have to be considered when deciding which data to use and how to analyse them, stemming from the explicit hypotheses being tested. Although this makes comparison between disparity analyses difficult and renders premature attempts to achieve the generalization required to answer the broad biological questions (e.g. how does phenotypic variation evolve?), this diversity of methodological approaches provides researchers with a great number of tools tailored to answer specific biological questions.

Many of the problems in morphological disparity analysis arise from 'blind' application of established methodological pipelines without consideration of the biological question being addressed. We advocate that researchers should assemble their analytical protocol based on an experimental approach that explores the impact of competing methods, such as choice of indices, ordination method and ancestral state estimation method on disparity analysis results. Thankfully, this is becoming easier through the availability of diverse, well-documented R packages for multidimensional analysis [42,65–68]. Many of the methods employed in disparity analysis are used more widely in other fields, including genomics and ecology, which also encompass analyses of multidimensional datasets [69–72]. Innovations in morphological disparity analyses likely await discovery in their respective literatures.

While studies of morphological disparity would benefit from advances in multidimensional analysis in other fields, the concept of a morphospace could reciprocally benefit other disciplines. For example, the multidimensional analysis of [47], which analysed patterns of form and function in plants, is essentially an eco-morphospace; isotopic analyses of organisms [52,73] can be represented as an isotope-space; ecosystem functioning in [69] as an ecosystem-space [74], etc. These generalizations could

also be exported for any set of traits: cognate approaches have been adopted in the analysis of single-cell comparative transcriptome data [75] where interpretation of the resulting transcriptome spaces would be improved by giving careful attention to the concerns we highlight concerning morphospaces.

Although disparity analyses are now simple to implement in freely available software [42,65–68], it is crucial to remember that they are multidimensional analyses and that multidimensional analyses are complex. We assert that future morphological analyses will benefit from emphasizing the methodological decisions made, rather than simply using disparity analysis because it exists.

Data accessibility. No data were generated in this paper.

Authors' contributions. T.G., N.C. and P.C.J.D. proposed this review; T.G. and N.C. led the writing supported by P.C.J.D. and G.H.T. All authors edited drafts and approved the final version.

Competing interests. We have no competing interests.

Funding. A.G. was funded by European Research Council Starting grant no. 637171 ADaPTiVE; A.L.J. was funded by an Irish Research Council Laureate Award IRCLA/2017/186; E.S. was funded by a University of Adelaide Research Fellowship; E.E.S. was funded by a Leverhulme Trust Research Project Grant (grant no. DGR01020); P.C.J.D. was funded by NERC (grant nos NE/P013678/1; NE/N002067/1) and BBSRC (grant no BB/N000919/1); S.L.B. was funded by European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program (grant agreement no. 756226, ERC Starting Grant: PalM) and a Leverhulme Trust Research Project Grant (grant no. RPG-2017-167); T.G. was funded by ARC DP170103227 and FT180100634 awarded to V.W.; G.H.T. was funded by European Research Council Consolidator grant no. 615709 ToLERates and Royal Society University Research Fellowship UF120016.

Acknowledgements. This article results from discussion at the Royal Society International Science Seminar on 'Reconciling disparate views on disparity' held at Chicheley Hall, 9–10 January 2018. We thank the two anonymous reviewers for their helpful and positive comments.

References

1. Runnegar B. 1987 Rates and modes of evolution in the Mollusca. In *Rates of evolution* (eds KSW Campbell, MF Day), pp. 39–60. London, UK: Allen & Unwin.
2. Gould SJ. 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. (doi:10.1017/S0094837300010745)
3. Mitteroecker P, Huttegger SM. 2009 The concept of morphospaces in evolutionary and developmental biology: mathematics and metaphors. *Biol. Theory* **4**, 54–67. (doi:10.1162/biot.2009.4.1.54)
4. Foote M. 1995 Morphological diversification of Paleozoic crinoids. *Paleobiology* **21**, 273–299. (doi:10.1017/S0094837300013300)
5. Briggs DE, Fortey RA, Wills MA. 1992 Morphological disparity in the Cambrian. *Science* **256**, 1670–1673. (doi:10.1126/science.256.5064.1670)
6. Raup DM. 1961 The geometry of coiling in gastropods. *Proc. Natl Acad. Sci. USA* **47**, 602–609. (doi:10.1073/pnas.47.4.602)
7. Gerber S. 2017 The geometry of morphospaces: lessons from the classic Raup shell coiling model. *Biol. Rev. Camb. Philos. Soc.* **92**, 1142–1155. (doi:10.1111/brv.12276)
8. Goswami A, Polly PD. 2010 The influence of modularity on cranial morphological disparity in Carnivora and Primates (Mammalia). *PLoS ONE* **5**, e9517. (doi:10.1371/journal.pone.0009517)
9. Bardua C, Wilkinson M, Gower DJ, Sherratt E, Goswami A. 2019 Morphological evolution and modularity of the caecilian skull. *BMC Evol. Biol.* **19**, 30. (doi:10.1186/s12862-018-1342-7)
10. Hipsley CA, Müller J. 2017 Developmental dynamics of ecomorphological convergence in a transcontinental lizard radiation. *Evolution* **71**, 936–948. (doi:10.1111/evo.13186)
11. Brusatte SL, Benton MJ, Ruta M, Lloyd GT. 2008 Superiority, competition, and opportunism in the evolutionary radiation of dinosaurs. *Science* **321**, 1485–1488. (doi:10.1126/science.1161833)
12. Prentice KC, Ruta M, Benton MJ. 2011 Evolution of morphological disparity in pterosaurs. *J. Syst. Paleontol.* **9**, 337–353. (doi:10.1080/14772019.2011.565081)
13. Guillerme T, Cooper N. 2018 Time for a rethink: time sub-sampling methods in disparity-through-time analyses. *Palaeontology* **61**, 481–493. (doi:10.1111/pala.12364)
14. Friedman M. 2010 Explosive morphological diversification of spiny-finned teleost fishes in the aftermath of the end-Cretaceous extinction. *Proc. R. Soc. B* **277**, 1675–1683. (doi:10.1098/rspb.2009.2177)
15. Anderson PSL, Friedman M. 2012 Using cladistic characters to predict functional variety: experiments using early gnathostomes. *J. Vertebr. Paleontol.* **32**, 1254–1270. (doi:10.1080/02724634.2012.694386)
16. Fortey RA, Briggs DEG, Wills MA. 1996 The Cambrian evolutionary 'explosion': decoupling cladogenesis from morphological disparity. *Biol. J. Linn. Soc. Lond.* **57**, 13–33. (doi:10.1111/j.1095-8312.1996.tb01693.x)
17. Hopkins MJ. 2013 Decoupling of taxonomic diversity and morphological disparity during decline of the Cambrian trilobite family Pterocephaliidae. *J. Evol. Biol.* **26**, 1665–1676. (doi:10.1111/jeb.12164)
18. Pierce SE, Angielczyk KD, Rayfield EJ. 2008 Patterns of morphospace occupation and mechanical performance in extant crocodilian skulls: a combined geometric morphometric and finite element modeling approach. *J. Morphol.* **269**, 840–864. (doi:10.1002/jmor.10627)
19. Wainwright PC, Alfaro ME, Bolnick DI, Hulsey CD. 2005 Many-to-one mapping of form to function: a general principle in organismal design? *Integr. Comp. Biol.* **45**, 256–262. (doi:10.1093/icb/45.2.256)
20. Tyler CL, Leighton LR. 2011 Detecting competition in the fossil record: support for character displacement among ordovician brachiopods. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **307**, 205–217. (doi:10.1016/j.palaeo.2011.05.020)
21. Foote M. 1997 The evolution of morphological diversity. *Annu. Rev. Ecol. Syst.* **28**, 129–152. (doi:10.1146/annurev.ecolsys.28.1.129)
22. Erwin DH. 2007 Disparity: morphological pattern and developmental context. *Palaeontology* **50**, 57–73. (doi:10.1111/j.1475-4983.2006.00614.x)
23. Hughes M, Gerber S, Wills MA. 2013 Clades reach highest morphological disparity early in their evolution. *Proc. Natl Acad. Sci. USA* **110**, 13 875–13 879. (doi:10.1073/pnas.1302642110)
24. Foote M. 1989 Perimeter-based Fourier analysis: a new morphometric method applied to the trilobite cranidium. *J. Paleontol.* **63**, 880–885. (doi:10.1017/S0022336000036556)
25. Deline B, Greenwood JM, Clark JW, Puttick MN, Peterson KJ, Donoghue PCJ. 2018 Evolution of metazoan morphological disparity. *Proc. Natl Acad. Sci. USA* **115**, E8909–E8918. (doi:10.1073/pnas.1810575115)
26. Sherratt E, Serb JM, Adams DC. 2017 Rates of morphological evolution, asymmetry and morphological integration of shell shape in scallops. *BMC Evol. Biol.* **17**, 248. (doi:10.1186/s12862-017-1098-5)
27. Spriggs EL, Schmerler SB, Edwards EJ, Donoghue MJ. 2018 Leaf form evolution in *Viburnum* parallels variation within individual plants. *Am. Nat.* **191**, 235–249. (doi:10.1086/695337)
28. Saunders WB, Work DM, Nikolaeva SV. 2004 The evolutionary history of shell geometry in Paleozoic ammonoids. *Paleobiology* **30**, 19–43. (doi:10.1666/0094-8373(2004)030<0019:TEHOSG>2.0.CO;2)
29. Hetherington AJ, Sherratt E, Ruta M, Wilkinson M, Deline B, Donoghue PC. 2015 Do cladistic and morphometric data capture common patterns of

morphological disparity? *Palaeontology* **58**, 393–399. (doi:10.1111/pala.12159)

30. Hopkins MJ, Gerber S. 2017 Morphological disparity. In *Evolutionary developmental biology* (eds L Nuño de la Rosa, GB Müller), pp. 1–12. New York, NY: Springer International Publishing.
31. Close RA, Friedman M, Lloyd GT, Benson RBJ. 2015 Evidence for a Mid-Jurassic adaptive radiation in mammals. *Curr. Biol.* **25**, 2137–2142. (doi:10.1016/j.cub.2015.06.047)
32. Palci A, Lee MSY. 2018 Geometric morphometrics, homology and cladistics: review and recommendations. *Cladistics* **35**, 230–242. (doi:10.1111/cla.12340)
33. Legendre P, Legendre LFJ. 2012 *Numerical ecology*. Amsterdam, The Netherlands: Elsevier.
34. Lehmann OER, Ezcurra MD, Butler RJ, Lloyd GT. 2019 Biases with the generalized Euclidean distance measure in disparity analyses with high levels of missing data. *Palaeontology* **62**, 837–849. (doi:10.1111/pala.12430)
35. Dryden IL, Mardia KV. 2016 *Statistical shape analysis: with applications in R*. Chichester, UK: John Wiley & Sons.
36. Bright JA, Marugán-Lobón J, Cobb SN, Rayfield EJ. 2016 The shapes of bird beaks are highly controlled by nondietary factors. *Proc. Natl Acad. Sci. USA* **113**, 5352–5357. (doi:10.1073/pnas.1602683113)
37. Bookstein FL. 2015 The relation between geometric morphometrics and functional morphology, as explored by Procrustes interpretation of individual shape measures pertinent to function. *Anat. Rec.* **298**, 314–327. (doi:10.1002/ar.23063)
38. Weisbecker V *et al*. 2019 Individual variation of the masticatory system dominates 3D skull shape in the herbivory-adapted marsupial wombats. *Front. Zool.* **16**, 41. (doi:10.1186/s12983-019-0338-5)
39. Gerber S. 2019 Use and misuse of discrete character data for morphospace and disparity analyses. *Palaeontology* **62**, 305–319. (doi:10.1111/pala.12407)
40. Cailliez F. 1983 The analytical solution of the additive constant problem. *Psychometrika.* **48**, 305–308. (doi:10.1007/BF02294026)
41. Hill JJ, Puttick MN, Stubbs TL, Rayfield EJ, Donoghue PCJ. 2018 Evolution of jaw disparity in fishes. *Palaeontology* **61**, 847–854. (doi:10.1111/pala.12371)
42. Lloyd GT. 2016 Estimating morphological diversity and tempo with discrete character-taxon matrices: implementation, challenges, progress, and future directions. *Biol. J. Linnean Soc.* **118**, 131–151. (doi:10.1111/bij.12746)
43. Lloyd GT. 2018 Journeys through discrete-character morphospace: synthesizing phylogeny, tempo, and disparity. *Palaeontology* **61**, 637–645. (doi:10.1111/pala.12380)
44. Gorman KB, Williams TD, Fraser WR. 2014 Ecological sexual dimorphism and environmental variability within a community of Antarctic penguins (genus Pygoscelis). *PLoS ONE* **9**, e90081. (doi:10.1371/journal.pone.0090081)
45. Horst A, Hill A, Gorman K. 2020 palmerpenguins v.0.0.0.9000. (https://allisonhorst.github.io/palmerpenguins/)
46. Guillerme T, Puttick MN, Marcy AE, Weisbecker V. 2020 Shifting spaces: which disparity or dissimilarity metrics best summarize occupancy in multidimensional spaces? *Ecol. Evol.* (doi:10.1002/ece3.6452)
47. Díaz S *et al*. 2016 The global spectrum of plant form and function. *Nature* **529**, 167–171. (doi:10.1038/nature16489)
48. Finlay S, Cooper N. 2015 Morphological diversity in tenrecs (Afrosoricida, Tenrecidae): comparing tenrec skull diversity to their closest relatives. *PeerJ* **3**, e927. (doi:10.7717/peerj.927)
49. Wills MA. 2001 Morphological disparity: a primer. In *Fossils, phylogeny, and form*, Topics in geobiology, vol. 19 (eds. JM Adrain, GD Edgecombe, BS Lieberman), pp. 55–144. Boston, MA: Springer.
50. Ciampaglio CN, Kemp M, McShea DW. 2001 Detecting changes in morphospace occupation patterns in the fossil record: characterization and analysis of measures of disparity. *Paleobiology* **27**, 695–715. (doi:10.1666/0094-8373(2001)027<0695:DCIMOP>2.0.CO;2)
51. Bellman R. 1966 Dynamic programming. *Science* **153**, 34–37. (doi:10.1126/science.153.3731.34)
52. Jackson AL, Inger R, Parnell AC, Bearhop S. 2011 Comparing isotopic niche widths among and within communities: SIBER—stable isotope Bayesian ellipses in R. *J. Anim. Ecol.* **80**, 595–602. (doi:10.1111/j.1365-2656.2011.01806.x)
53. Clavel J, Escarguel G, Merceron G. 2015 mvMORPH: an R package for fitting multivariate evolutionary models to morphometric data. *Methods Ecol. Evol.* **6**, 1311–1319. (doi:10.1111/2041-210X.12420)
54. Adams DC, Collyer ML. 2018 Multivariate phylogenetic comparative methods: evaluations, comparisons, and recommendations. *Syst. Biol.* **67**, 14–31. (doi:10.1093/sysbio/syx055)
55. Anderson MJ. 2001 A new method for non-parametric multivariate analysis of variance. *Austral. Ecol.* **26**, 32–46. (doi:10.1046/j.1442-9993.2001.01070.x)
56. Collyer ML, Sekora DJ, Adams DC. 2015 A method for analysis of phenotypic change for phenotypes described by high-dimensional data. *Heredity* **115**, 357–365. (doi:10.1038/hdy.2014.75)
57. Adams DC. 2014 A method for assessing phylogenetic least squares models for shape and other high-dimensional multivariate data. *Evolution* **68**, 2675–2688. (doi:10.1111/evo.12463)
58. Strube MJ. 1988 Bootstrap type I error rates for the correlation coefficient: an examination of alternate procedures. *Psychol. Bull.* **104**, 290. (doi:10.1037/0033-2909.104.2.290)
59. Harvey PH, Pagel MD. 1998 *The comparative method in evolutionary biology*. Oxford, UK: Oxford University Press.
60. Uyeda JC, Caetano DS, Pennell MW. 2015 Comparative analysis of principal components can be misleading. *Syst. Biol.* **64**, 677–689. (doi:10.1093/sysbio/syv019)
61. Gerber S. 2014 Not all roads can be taken: development induces anisotropic accessibility in morphospace. *Evol. Dev.* **16**, 373–381. (doi:10.1111/ede.12098)
62. Blomberg SP, Rathnayake SI, Moreau CM. 2020 Beyond Brownian motion and the Ornstein-Uhlenbeck process: stochastic diffusion models for the evolution of quantitative characters. *Am. Nat.* **195**, 145–165. (doi:10.1086/706339)
63. Brusatte SL, Montanari S, Yi H-y, Norell MA. 2011 Phylogenetic corrections for morphological disparity analysis: new methodology and case studies. *Paleobiology* **37**, 1–22. (doi:10.1666/09057.1)
64. Louca S, Pennell MW. 2020 Extant timetrees are consistent with a myriad of diversification histories. *Nature* **580**, 502–505. (doi:10.1038/s41586-020-2176-1)
65. Bouxin G. 2005 Ginkgo, a multivariate analysis package. *J. Veg. Sci.* **16**, 355–359. (doi:10.1111/j.1654-1103.2005.tb02374.x)
66. Oksanen J, Kindt R, Legendre P, O'Hara B, Stevens MHH, Oksanen MJ, Suggests M. 2007 The vegan package. *Commun. Ecol. Pack.* **10**, 631–637.
67. Harmon LJ, Weir JT, Brock CD, Glor RE, Challenger W. 2008 GEIGER: investigating evolutionary radiations. *Bioinformatics* **24**, 129–131. (doi:10.1093/bioinformatics/btm538)
68. Guillerme T. 2018 dispRity: a modular R package for measuring disparity. *Methods Ecol. Evol.* **9**, 1755–1763. (doi:10.1111/2041-210X.13022)
69. Donohue I *et al*. 2013 On the dimensionality of ecological stability. *Ecol. Lett.* **16**, 421–429. (doi:10.1111/ele.12086)
70. Saupe EE, Qiao H, Hendricks JR, Portell RW, Hunter SJ, Soberón J, Lieberman BS. 2015 Niche breadth and geographic range size as determinants of species survival on geological time scales. *Global Ecol. Biogeogr.* **24**, 1159–1169. (doi:10.1111/geb.12333)
71. Canter EJ, Cuellar-Gempeler C, Pastore AI, Miller TE, Mason OU. 2018 Predator identity more than predator richness structures aquatic microbial assemblages in *Sarracenia purpurea* leaves. *Ecology* **99**, 652–660. (doi:10.1002/ecy.2128)
72. Mammola S. 2019 Assessing similarity of *n*-dimensional hypervolumes: which metric to use? *J. Biogeogr.* **46**, 2012–2023. (doi:10.1111/jbi.13618)
73. Swanson HK, Lysy M, Power M, Stasko AD, Johnson JD, Reist JD. 2015 A new probabilistic method for quantifying *n*-dimensional ecological niches and niche overlap. *Ecology* **96**, 318–324. (doi:10.1890/14-0235.1)
74. Qiao H, Escobar LE, Saupe EE, Ji L, Soberón J. 2017 A cautionary note on the use of hypervolume kernel density estimators in ecological niche modelling. *Global Ecol. Biogeogr.* **26**, 1066–1070. (doi:10.1111/geb.12492)
75. Sebé-Pedrós A *et al*. 2018 Early metazoan cell type diversity and the evolution of multicellular gene regulation. *Nat. Ecol. Evol.* **2**, 1176–1188. (doi:10.1038/s41559-018-0575-6)