UNIVERSITY OF LEEDS

This is a repository copy of *Journeys through discrete-character morphospace: synthesizing phylogeny, tempo, and disparity*.

White Rose Research Online URL for this paper:
http://eprints.whiterose.ac.uk/136318/

Version: Accepted Version

Article:

Lloyd, GT orcid.org/0000-0001-6887-3981 (2018) Journeys through discrete-character morphospace: synthesizing phylogeny, tempo, and disparity. Palaeontology, 61 (5). pp. 637-645. ISSN 0031-0239

https://doi.org/10.1111/pala.12380

© The Palaeontological Association. This is the peer reviewed version of the following article: Lloyd, GT (2018) Journeys through discrete-character morphospace: synthesizing phylogeny, tempo, and disparity. Palaeontology, 61 (5). pp. 637-645. ISSN 0031-0239, which has been published in final form at https://doi.org/10.1111/pala.12380. This article may be used for non-commercial purposes in accordance with Wiley Terms and Conditions for Use of Self-Archived Versions.

Reuse
Items deposited in White Rose Research Online are protected by copyright, with all rights reserved unless indicated otherwise. They may be downloaded and/or printed for private study, or other acts as permitted by national copyright laws. The publisher or other rights holders may allow further reproduction and re-use of the full text version. This is indicated by the licence information on the White Rose Research Online record for the item.

Takedown
If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing eprints@whiterose.ac.uk including the URL of the record and the reason for the withdrawal request.

White Rose
university consortium
Universities of Leeds, Sheffield & York

eprints@whiterose.ac.uk
https://eprints.whiterose.ac.uk/

Journeys through discrete-character morphospace: synthesising phylogeny, tempo, and disparity

Graeme T. Lloyd[1]

[1]School of Earth and Environment, Maths/Earth and Environment Building, University of Leeds, Leeds, LS2 9JT, UK

Abstract: Palaeontologists have long employed discrete categorical data to capture morphological variation in fossil species, using the resulting character-taxon matrices to measure evolutionary tempo, infer phylogenies, and capture morphological disparity. However, to date these have been seen as separate approaches despite a common goal of understanding morphological evolution over deep time. Here I argue that there are clear advantages to considering these three lines of enquiry in a single space: the phylomorphospace. Conceptually these high-dimensional spaces capture how a phylogenetic tree explores morphospace and allow us to consider important process questions around evolutionary rates, constraints, convergence, and directional trends. Currently the literature contains fundamentally different approaches used to generate such spaces, with no direct comparison between them despite the differing evolutionary histories they imply. Here I directly compare five different phylomorphospace approaches, three with direct literature equivalents and two that are novel. I use a single empirical case study of coelurosaurian theropod dinosaurs (152 taxa, 853 characters) to show that under many analyses the literature-derived approaches tend to reflect introduced phylogenetic – rather than the intended morphological – signal. The two novel approaches, which produce limited ancestral state estimates prior to ordination, are able to minimise this phylogenetic signal and thus exhibit more realistic amounts of phylogenetic signal, rate heterogeneity, and convergent evolution.

Key words [6 max]: ancestors; cladistics; disparity; morphology; phylogeny; tempo

Introduction

Morphological variation has been captured as discrete categorical data since at least the 1940s. Such character-taxon matrices were first inspired as an extension of Simpson's (1944) Tempo and Mode, with the intent to capture morphological change across the whole organism (Olson 1944; Westoll 1949). Subsequently they were employed in systematics to infer evolutionary relationships (Sokal and Sneath 1963; Sneath and Sokal 1973), and most recently they were utilised to measure morphological diversity, or disparity (Foote 1991). However, despite sharing a common goal of understanding morphological evolution these three main strands – tempo, phylogeny, and disparity – have been treated as separate enterprises (e.g., as in Brusatte et al 2014). However, there is potential to gain additional insights by considering these three strands collectively.

An initial challenge to this endeavour is establishing a framework where tempo, phylogeny, and disparity can be considered simultaneously. However, a clear solution presents itself in the form of the "phylomorphospace" (first formally named by Sidlauskas 2008, but conceptually traceable back to Stone 2003). Such spaces are extensions of morphospaces, usually high-dimensional (multivariate) ordination spaces with axes that, often only abstractly (but see Wright 2017), represent overall morphological variation: points that are proximal in the space are morphologically similar and vice versa. The phylo- prefix reflects the projection of estimated ancestors and plotting of branches into the same space (Stone 2003). Such spaces can also be thought of as containing “stations” (nodes of a phylogenetic tree) connected by “journeys” (branches of the tree), with each root-to-tip pathway representing the shortest journey through morphospace that led to a specific tip morphology (Figure 1).

A critical advantage of such spaces is they can enable the switch from simply establishing pattern (overall distribution of points in a space), to considering process (the direction morphological evolution has taken between those points). For example, previous authors have shown how such spaces can be used to help differentiate between evolutionary rates and constraints (Sidlauskas 2008), expose convergent evolution (Stayton 2015; Page and Cooper 2017), or reveal directional biases in trends in morphospace exploration between subclades (Hopkins 2016).

Discrete characters represent a unique challenge for phylomorphospace construction. Many of these were previously summarised by Lloyd (2016) and are pertinent to generating non-phylogenetic morphospaces from discrete data (e.g., projecting non-Euclidean distances into a Euclidean space, and adequately visualising the data when the variance is often spread over a high number of axes). However, here I will focus on a single unexplored issue, namely that there are two fundamentally different ways to project estimated ancestral morphologies into such spaces (Figure 2). These are here termed post-, and pre-ordination ancestral state estimation (post-OASE and pre-OASE hereafter). Post-OASE is the typical form (e.g., Hopkins and Smith 2015, Wright 2017) and occurs across all types of morphospaces, with continuous ancestral states being estimated directly from the ordination axes. Pre-OASE is less common and is possibly unique to discrete characters (e.g., Brusatte et al 2011; Butler et al 2012). Here ancestral states are estimated prior to ordination by generating sets of discrete states for each node in the tree (see methods and Figure 3). These two approaches necessarily lead to fundamentally different phylomorphospaces and hence have a major effect on any inferences about evolutionary processes made.

Here I conduct the first, to my knowledge, direct comparison of post- and pre-OASE phylomorphospaces using a single empirical case study that was initially conceived to examine tempo, phylogeny, and disparity separately (Brusatte et al 2014). Although a more comprehensive set of comparisons across multiple data sets would clearly be desirable, current implementations of the phylomorphospace algorithms are too slow to permit this. Nevertheless, the protocols outlined here allow future workers to examine the same effects in their own data. Here I particularly focus on the degree to which the resulting phylomorphospaces represent phylogenetic rather than morphological variation (i.e., any two tips must continuously diverge from their common ancestor, increasing their phylogenetic distance). This concern is motivated by a desire to avoid erroneous inferences about evolution. For example, phylogenetic variation will necessarily be continuously divergent, whereas morphological evolution is known to be at least partially convergent. Similarly, phylogenetic signal represents homogenous evolutionary rates, whereas morphological evolution is widely considered heterogeneous. Thus overreliance on phylogeny may lead us to misconstrue evolutionary processes.

Alternative optimality criteria are of course worth considering, but will introduce their own complexities. For example, using the discrete character simulation approaches other workers have applied to phylogenetic inference (e.g., Wright and Hillis 2014; O'Reilly et al 2016) would necessarily introduce phylogenetic signal a priori. It is hoped solutions to these issues may be discovered in future, but minimally this study shows empirically that major differences can arise between phylomorphospace approaches even when applied to the same input data.

Materials and Methods

The discrete-character taxon matrix used here is derived directly from Brusatte et al (2014), and contains a mixture of binary, ordered, and unordered multistate characters as well as polymorphisms. As the implementations applied here are slow (taking several days to run all analyses on a standard laptop), the principal aim is not to understand phylogenetic uncertainty a single phylogenetic hypothesis was used. This represents the first most parsimonious tree recovered from the reanalysed version of the matrix available at graemetlloyd.com/matrdino.html. This was time-scaled using tip dates from the Paleobiology Database and the timePaleoPhy function of the paleotree Package (Bapst 2012).

Six separate phylomorphospaces were generated from the data. These include the four pre-OASE methods (Figure 3), a single post-OASE method, and a control "phylospace", where phylogenetic (in millions of years) rather than morphological distances were ordinated. In all cases the phylomorphospace was generated by performing a principal coordinate (Gower 1966) ordination using the cmdscale function in base R (R Core Team 2017). The four pre-OASE methods are all based on a likelihood ancestral state method (Yang et al 1995) and represent the range of possible outcomes from two binary options (Figure 3; the AncStateEstMatrix function in Claddis, Lloyd 2016). These represent choices to estimate missing or uncertain values using a phylogenetic hypothesis and thus are likely to represent varying degrees of phylogenetic signal. Specifically, in pre-OASE1 the only output is ancestral state estimates for internal nodes that have direct descendants with non-missing (or non-inapplicable) states. For pre-OASE2 the only additional output is a collapsing of uncertain (polymorphic) tip values. Under pre-OASE3 no tip outputs are produced, but ancestral state estimates are made for all internal nodes, even if their direct descendants have missing (or inapplicable) states. Finally, pre-OASE4 returns output for every internal node and every tip. The post-OASE method simply applies the ace (ancestral character estimation) function in the ape package (Paradis et al 2004) using the principal coordinate axes themselves as input.

Five different analyses were performed to both qualitatively and quantitatively assess and compare the six phylomorphospaces. Specific analyses performed were: 1) scree plots (summarising the distribution of the variance over the ordination axes), 2) simple bivariate phylomorphospace plots of the first two ordination axes (allowing only a cursory inspection of the data due to the low variances of these axes), 3) rate heterogeneity plots (branch length against branch duration), 4) pairwise convergence histograms applying the C_1 metric of Stayton (2015), and 5) correlations between phylogenetic and ordination space distances between taxa on the time-scaled tree. All analyses were performed in R (R Core Team 2017) with the full script available in Lloyd (2018).

Results

The main results are summarised in Figure 4 and Table 1.

Scree plots (Figure 4A-F) exhibit the general issue with discrete-character ordination spaces of variance being dispersed over a large number of axes (Lloyd 2016), with low variance on the first two diminishing the value of the bivariate plots (Figure 4G-L). However, Figure 4E is of note as it includes a second deflection point (at around principal coordinate axis 100). This is the result of a correction for negative eigenvalues (Caillez 1983), but is unique here to the post-OASE approach. Thus including the phylogeny directly in the ordination (pre-OASE; Figure 4A-D) has a clear advantage in generating ordinations that require less distortion when moving from raw (non-Euclidean) morphological distances to ordinated (Euclidean) morphological distances.

Ideal phylomorphospaces should exhibit a good spread of points (as the pairwise distances should themselves be normally distributed). However, it is clear that one of the pre-OASE approaches (pre-OASE4; Figure 4J) has an unusually clustered, strongly V-shaped distribution that is more comparable to the "phylospace" plot (Figure 4L) than the other phylomorphospaces (Figure 4G-I,K). These results should lead us to be extremely cautious about going too far in using phylogenies to "correct" for missing data, and to particularly avoid the pre-OASE4 option.

Rate heterogeneity is captured in Figure 4M-R by plotting branch lengths (Euclidean distance in the ordination space) against branch duration (in millions of years). The expectation here under 100% phylogenetic signal is that all branches exhibit the same mean rate (homogenous rates) and hence fall on the corresponding dashed line. Here we see that this is not the case even for the phylospace (Figure 4R). However, all pre-OASE plots (Figure 4M-P) show considerably larger rate heterogeneity, even having some branches falling outside the heat map area (over one hundred times the mean rate, or below one-hundredth the mean rate). By contrast post-OASE rates are considerably more homogenous, more closely reflecting the phylogenetic signal (Table 1).

Convergent evolution was measured for all pairwise tip comparisons using the C_1 metric of Stayton (2015), being one minus the Euclidean distance between the two tips over the largest distance achieved by their lineages since they diverged from their most recent common ancestor (Figure 4S-X). The expectation here is that under pure phylogenetic signal all values will be zero, indicating the largest divergence was achieved at the tips. Again, this is not quite the case (Figure 4X; Table 1) due to the distorting of non-Euclidean values into a Euclidean space, although most values are either zero, or very close to zero. By contrast under pre-OASE approaches a clear tail of higher C_1 values, indicating some degree of convergence, can be seen (Figure 4S-C), along with a reduction in the size of this tail as greater levels of missing values are phylogenetically predicted. However, most notable is the post-OASE approach that shows not just extremely low convergence but even lower convergence than pure phylogenetic signal. Here this is considered extremely implausible, especially as other character-based metrics (e.g., see Hoyal Cuthill et al 2010) would comfortably show some degree of convergence in the data.

Finally, the overall degree to which phylogeny controls the phylomorphospace generated was assessed using a simple Pearson's correlation of the Euclidean distance between tips in the ordination space and the phylogenetic distances (in millions of years) between tips on the tree (Table 1). A high value indicates a strong phylogenetic signal, whereas a low value suggests a stronger potential morphological signal. Here I consider the latter ideal, but note that we would logically expect some phylogenetic signal in morphology. Additionally, because the expectation of 100% correlation in the phylospace is not met (again due to the non-Euclidean nature of phylogenetic distances) all values were rescaled against the phylospace value to return an estimate of percent phylogenetic signal (Table 1). Thus roughly half of the post-OASE signal can be explained by phylogeny, whereas the pre-OASE approaches vary dramatically, from roughly one-quarter to almost three-quarters phylogenetic signal. That the lowest value can be reached by the pre-OASE1 approach further supports this as the optimal, i.e., least biased by introduced phylogenetic signal, phylomorphospace approach.

Discussion

At face value the post-OASE approach should be the ideal way to generate phylomorphospaces. It is certainly the most common approach in the literature as it is directly applicable regardless of the type of morphological data or the type of ordination employed (e.g., Page and Cooper 2017; Sherratt et al 2017). In its implementation it is usually faster, as it is applied after the dimensionality of the data has been reduced by ordination, and it has the theoretical advantage of allowing phylogenetic uncertainty to be expressed visually in the same space (as changing the tree does not "move" the tip values in the ordination), although I am not aware of this being done. However, there are some intuitive causes for concern too. For example, ancestral values will necessarily fall within the range of the sampled tip values (Stayton 2015), forcing the implicit assumptions that our sample already includes the morphological extremes and that ancestral values are always morphologically average. Furthermore, ancestral states estimated directly from the ordination axes are not mappable back to a tangible discrete morphology diminishing their utility. Here these concerns are added to by showing that some simple interrogations of a post-OASE phylomorphospace exhibit patterns better explained directly by the phylogeny used to generate it than the morphological signal we are interested in. These include an implausibly low amount of convergent evolution and generally low rate heterogeneity.

By contrast pre-OASE approaches have some clear limitations. They are more convoluted; requiring more complex ancestral state estimations, and usually more of them: the number of characters tends to be greater than the number of taxa and hence, by mathematical necessity, the number of ordination axes. Additionally, any changes to the phylogenetic hypothesis require new ancestral state estimates and hence the generation of a new phylomorphospace, adding further computation time. However, despite these limitations they have still been favoured before, primarily by palaeontologists as a means of increasing sample size (Brusatte et al 2011) or addressing severe missing data levels (Butler et al 2012). Other advantages include removing the assumption that estimated ancestors must fall within the range of the tips (compare Hopkins and Smith 2015, their Figure 3 with Brusatte et al 2011, their Figure 3). However, as shown here, there is considerable variation in implied evolutionary history amongst the four pre-OASE approaches. The results obtained here suggest those currently adopted in the literature (pre-OASE3-4) are likely suboptimal, introducing substantial phylogenetic signal that may be overwriting the true morphological pattern and generating peculiar phylomorphospaces (Figure 4J) as well as low amounts of convergent evolution and rate heterogeneity. However, this need not be the case, with already available options (pre-OASE1-2) in Claddis (Lloyd 2016) that can: 1) minimise this introduced signal, by over 45% in the example data set used here, 2) more faithfully retain the uncertainty of the empirical observations rather than "diluting" them with predicted values, and 3) avoid nonsensical issues like estimating either a single state for a truly polymorphic character or any value for an inapplicable character.

Minimally this study shows that different approaches lead to very different phylomorphospaces (and hence inferred evolutionary histories), but it does not directly address their utility in general. Previously I have argued for an apparently contradictory position: that there are good reasons to avoid ordination entirely (Lloyd 2016). Some of those reasons are evident here, particularly the distortions associated with projecting non-Euclidean distances into Euclidean ordination spaces. We might ask, then, whether the advantages of a phylomorphospace can still be enjoyed without the problems associated with ordination? This can be done by examining some of the phylomorphospace approaches introduced by previous authors (Sidaluskas 2008; Stayton 2015; Hopkins 2016). The methods of Sidlauskas (2008) require both branch-lengths and clade volumes, as specified by the ordination space. However, branch-lengths can be estimated outside of an ordination space and non-ordination disparity metrics can be employed as volume proxies (Lloyd 2016). Stayton's (2015) convergence metrics similarly rely on distances from an ordination space, but these can certainly be estimated without ordination and indeed would arguably be superior as they would avoid distorting these distances. However, the directionality bias measures used by Hopkins (2016) do not provide a clear non-ordination alternative. Significantly, though, any ordination-free alternative to generating a phylomorphospace requires ancestral state estimation to be made prior to (i.e., without) ordination. Consequently, only the pre-OASE approaches even allow for these options, further supporting their use over the post-OASE approach.

This study joins several others in emphasising the immaturity of our understanding of discrete character evolution. For example, Lloyd (2016) showed that a novel distance metric was in many cases superior to those previously applied in disparity studies. (Directly relevant here as a step in the phylomorphospace pipeline, Figure 2.) Hoyal Cuthill (2015a,b) extended earlier homoplasy metrics (Hoyal Cuthill 2010) to a more comprehensive understanding of how evolution explores the "state space" of discrete characters. This is related to another important concept in discrete character evolution, specifically the rapid "exhaustion" of novel states that appears to be a common pattern across multiple clades (Wagner 2000). This is logically related to the limits of disparity (Oyston et al 2015), and hence is directly relevant to our understanding of phylomorphospaces. For example, these phenomena might serve to explain some distance-from-root patterns generated in an earlier version of this manuscript, where the root morphology is rapidly left behind with subsequent evolution apparently constrained at the "edges" of the hyperdimensional space (see Lloyd 2018). Collectively, these studies show that despite almost 75 years of assembling character-taxon matrices there is much to be revealed from studying morphology in the form of discrete categorical data and further foundational changes can be expected.

Conclusion

Phylomorphospaces can allow us to consider phylogeny, tempo, and mode simultaneously when inferring morphological evolution from discrete characters. However, different approaches exist to generate such spaces and these can (and do) lead to fundamentally different interpretations of evolutionary history. The results presented here suggest caution be applied when using all of the most common approaches from the literature – whether ancestral states are estimated pre- or post-ordination. More generally, the fact that so much of the result can reflect phylogenetic rather than morphological signal should urge us to be cautious whenever we use phylogeny to “correct” the fossil record. The results shown here also highlight issues related to distortions generated by ordinating non-Euclidean distances into Euclidean spaces and future research should concentrate on developing ordination-free versions of phylomorphospace analyses. Finally, that this particular comparison has not been explored previously further emphasises the immature nature of these methods and the potential for far-reaching foundational improvements to our understanding of morphological evolution as captured by discrete categorical data.

Acknowledgements

For discussions and help with R code, both for this study and the development of the Claddis package in general, I would like to thank: Dave Bapst, Mark Bell, Roger Benson, Steve Brusatte, Rich Fitzjohn, Dani Fraser, Matt Friedman, Thomas Guillerme, Melanie Hopkins, Matt Pennell, Liam Revell, Emma Sherratt, Graham Slater, Pete Smits, and Steve Wang. For funding across the lifespan of the Claddis package (and the seeds of this study) I would like to thank the Natural Environment Research Council (Grant NE/I005536/1 to Matt Friedman), the Australian Research Council (Grant DE140101879 to GTL), and both the University of Oxford and Macquarie University. This manuscript was improved by thoughtful reviews by Richard Butler and Jen Hoyal Cuthill.

Data archiving statement

Data for this study are available in the Dryad Digital Repository: http://datadryad.org/review?doi=doi:10.5061/dryad.1j5h875 [please note that the data for this paper are not yet published and this temporary link should not be shared without the express permission of the author].

References

Bapst, D. W., 2012. paleotree: an R package for paleontological and phylogenetic analyses of evolution. Methods in Ecology and Evolution. **3,** 803-807.

Brusatte, S. L., 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.

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 dinosaur-bird transition. Current Biology, **24**, 2386-2392.

Butler, R. J., Brusatte, S. L., Andres, B. and Benson, R. B. J., 2012. How do geological sampling biases affect studies of morphological evolution in deep time? A case study of Pterosauria (Reptilia: Archosauria) disparity. Evolution, **66**, 147-162.

Cailliez, F., 1983. The analytical solution of the additive constant problem. Psychometrika, **48**, 305–308.

Foote, M., 1991. Morphological and taxonomic diversity in a clade's history: the blastoid record and stochastic simulations. Contributions from the Museum of Paleontology, The University of Michigan, **28**, 101-140.

Gower, J. C., 1966. Some distance properties of latent root and vector methods used in multivariate analysis. Biometrika, **53**, 325-338.

Hopkins, M. J., 2016. Magnitude versus direction of change and the contribution of macroevolutionary trends to morphological disparity. Biological Journal of the Linnean Society, **118**, 116-130.

Hopkins, M. J. and Smith, A. B., 2015. Dynamic evolutionary change in post-Paleozoic echinoids and the importance of scale when interpreting changes in rates of evolution. PNAS, **112**, 3758–3763.

Hoyal Cuthill, J. F., 2015a. The size of the character state space affects the occurrence and detection of homoplasy: modelling the probability of incompatibility for unordered phylogenetic characters. Journal of Theoretical Biology, **366**, 24-32.

Hoyal Cuthill, J. F., 2015b. The morphological state space revisited: what do phylogenetic patterns in homoplasy tell us about the number of possible character states? Interface Focus, **5**, 20150049.

Hoyal Cuthill, J. F., Braddy, S. J. and Donoghue, P. C. J., 2010. A formula for maximum possible steps in multistate characters: isolating matrix parameter effects on measures of evolutionary convergence. Cladistics, **26**, 98-102.

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.

Lloyd, G. T., 2018. Data from: Journeys through discrete-character morphospace: synthesising phylogeny, tempo, and disparity.

O'Reilly, J. E., Puttick, M. N., Parry, L., Tanner, A. R., Tarver, J. E., Fleming, J., Pisani, D. and Donoghue, P. C. J., 2016. Bayesian methods outperform parsimony but at the expense of precision in the estimation of phylogeny from discrete morphological data. Biology Letters, **12**, 20160081.

Olson, E. C., 1944. Origin of mammals based upon cranial morphology of the therapsid suborders. Geological Society of America Special Papers, **55**, 1-136.

Oyston, J. W., Hughes, M., Wagner, P. J., Gerber, S. and Wills, M. A., 2015. What limits the morphological disparity of clades? Interface Focus, **5**, 20150042.

Page, C. E. and Cooper, N. 2017. Morphological convergence in 'river dolphin' skulls. PeerJ, **5**, e4090.

Paradis, E., Claude, J. and Strimmer, K., 2004. APE: analyses of phylogenetics and evolution in R language. Bioinformatics, **20**, 289-290.

R Core Team, 2017. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Sherratt, E., Vidal-Garcia, M., Anstis, M. and Keogh, J. S., 2017. Adult frogs and tadpoles have different macroevolutionary patterns across the Australian continent. Nature Ecology & Evolution, **1**, 1385-1391.

Sidlauskas, B., 2008. Continuous and arrested morphological diversification in sister clades of characiform fishes: a phylomorphospace approach. Evolution, **62**, 3135-3156.

Sneath, P. H. A. and Sokal, R. R., 1973. Numerical Taxonomy. W. H. Freeman and Company, San Francisco, pp573.

Sokal, R. R. and Sneath, P. H. A., 1963. Principles of Numerical Taxonomy. W. H. Freeman and Company, San Francisco, pp359.

Stayton, C. T., 2015. The definition, recognition, and interpretation of convergent evolution, and two new measures for quantifying and assessing the significance of convergence. Evolution, **69**, 2140-2153.

Stone, J. R., 2003. Mapping cladograms into morphospaces. Acta Zoologica, **84**, 63-68.

Wagner, P. J., 2000. Exhaustion of morphologic character states among fossil taxa. Evolution, **54**, 365-386

Westoll, T. S., 1949. On the evolution of the Dipnoi. In: Jepsen, G. L., Simpson, G. G. and Mayr, E. (eds.), Genetics, Paleontology and Evolution. Princeton University Press, Princeton, p121–184.

Wright, D. F., 2017. Phenotypic innovation and adaptive constraints in the evolutionary radiation of Palaeozoic crinoids. Scientific Reports, **7**, 13745.

Wright, A. M. and Hillis, D. M., 2014. Bayesian analysis using a simple likelihood model outperforms parsimony for estimation of phylogeny from discrete morphological data. PLoS ONE, **9**, e109210.

Yang, Z., Kumar, S. and Nei, M., 1995. A new method of inference of ancestral nucleotide and amino acid sequences. Genetics, **141**, 1641-1650.

Figure 1 – A hypothetical phylomorphospace (redrawn from Sidlauskas 2008). Such spaces are typically high-dimensional and thus can never be fully visualised. Here, as is typical, only the first two axes (those containing the most variation) are displayed. Large black circles denote the tip values that are always used to infer the ordination (i.e., morpho-) space. Smaller grey circles represent estimated ancestral values that may be estimated either prior to (pre-OASE; Brusatte et al 2011), or after the ordination (post-OASE; Stone 2003). Grey lines denote branches of the phylogenetic hypothesis used, thus the phylomorphospace explicitly contains the phylogeny. Disparity (morphological diversity) is reflected in the spread of points in the space, and rates can be inferred from the length of the branches in the space (i.e., these represent the rate numerator). Thus a phylomorphospace allows us to consider phylogeny, disparity, and tempo simultaneously.

Figure 2 – Schematic of the two main routes (pipelines) to generating a discrete-character phylomorphospace (modified from Lloyd 2016). The primary difference is

whether the phylogenetic hypothesis is introduced prior to (pre-OASE), or after (post-OASE), ordination (dashed vertical line). If introduced prior to ordination, then ancestral estimates are for the discrete characters themselves (and missing tip values can also be estimated – see Figure 4). If introduced after ordination, then ancestral estimates are for the ordination axes (continuous data).

Figure 3 – Schematic showing four possible pre-Ordination Ancestral State Estimates (pre-OASE1-4) that result from two binary questions: whether to estimate tips as well as nodes (i.e., to fill missing values or collapse polymorphisms) and whether to estimate all states or just those for which tip data (i.e., direct descendants of that node) are available. An example input tree with six tips (labelled 1-6) and five internal nodes (A-E) is shown at upper left for a single binary character (white or black) with a polymorphism for tip 3 and missing data for tips 2 and 4. (Note that missing data is replaced with polymorphisms if estimating all states.) Trees on grey backgrounds represent the underlying marginal likelihoods estimated when using the associated AncStateEstMatrix function in the Claddis package (Lloyd 2016; github.com/graemetlloyd/Claddis), but note that conceptually the same general kinds of estimates could be made using other implementations or optimality criteria (e.g., parsimony, Bayesian). Finally, the lower part of the plot shows the output of estimated tip and node states (Claddis collapses marginal likelihoods to the most likely state, or returns a polymorphism if state likelihoods are exactly equal). Note that the four approaches are numbered from least (pre-OASE1) to most (pre-OASE4) missing values estimated, which can also be thought of as least to most introduced phylogenetic signal. Pre-OASE approaches in the literature correspond to pre-OASE3 (Brusatte et al 2011) or pre-OASE4 (Butler et al 2012), whereas pre-OASE1 and 2 are novel. Additionally, if no polymorphisms are present in the input data then pre-OASE1 and pre-OASE2 are identical.

Figure 4 – Primary results comparing six different phylomorphospace approaches applied to the same discrete character-taxon matrix and phylogenetic hypothesis (Brusatte et al 2014). Each "row" of results, from top to bottom, corresponds to: pre-OASE1-4, post-OASE, and the "phylospace" control. Each "column" of results, from left to right, corresponds to: 1) scree plots (the distribution of variance across ordination axes; A-F), 2) bivariate phylomorphospace plots (G-L; nodes shown as red circles and branches as grey lines), 3) bivariate plots of branch duration (x-axis, logged) and branch length (M-R; points represent individual branches of the tree with heat maps running from one-hundred (top, left), to one-hundredth (bottom, right), times the mean rate, indicated by the dashed red line), and 4) histograms of pairwise convergence values (C_1 metric of Stayton 2015). See text for interpretation of results.

Lay statement: Palaeontologists are interested in the diversity and tempo of evolutionary change along the branches of the tree of life, but these are typically treated as separate numerical analyses. Here I argue for a statistical technique that allows us to treat these aspects of evolution simultaneously, allowing a more complete understanding of how evolution has explored the possibilities of physical form. Specifically, I compare five different approaches, two novel to this study, showing that the three pre-existing approaches may contain biases that stem from a different signal (evolutionary relationships) than we intend to capture (the exploration of physical form).

Table 1 – Comparative summary statistics of the six phylomorphospace approaches applied to the empirical case study (Brusatte et al 2014). Phylogenetic signal was measured by correlating (Pearson's r; data were normally distributed) the phylogenetic distances between tips (in millions of years) with the distance between those tips in the complete ordination space (i.e., using all axes). Rate homogeneity was measured by correlating (Kendall's tau; data had a clear skew and contained ties) branch lengths in the complete ordination space with branch durations (in millions of years). In both cases the raw data were rescaled by dividing through by the phylospace value and converting to percentages. Convergent evolution was measured using the C_1 metric of Stayton (2015) across all pairwise tip-to-tip comparisons and using the complete ordination space (i.e., all axes).

Phylomorphospace approach	**Phylogenetic signal**		**Rate homogeneity**		**Convergent evolution**
	Raw	Rescaled	Raw	Rescaled	Mean pairwise C_1
Pre-OASE1	0.210	25.0%	0.217	26.3%	0.075
Pre-OASE2	0.224	26.5%	0.221	26.8%	0.074
Pre-OASE3	0.357	42.3%	0.392	47.6%	0.047
Pre-OASE4	0.604	71.7%	0.357	43.3%	0.028
Post-OASE	0.432	51.2%	0.513	62.3%	0.000
Phylospace	0.843	100.0%	0.824	100.0%	0.012