



DISCUSSION

REPLY TO COMMENTS ON: MACROEVOLUTIONARY PATTERNS IN RHYNCHOCEPHALIA: IS THE TUATARA (*SPHENODON PUNCTATUS*) A LIVING FOSSIL?

by JORGE A. HERRERA-FLORES , THOMAS L. STUBBS  and MICHAEL J. BENTON* 

School of Earth Sciences, University of Bristol, Wills Memorial Building, Queens Road, Bristol, BS8 1RJ, UK; jorge.herreraflores@bristol.ac.uk, tom.stubbs@bristol.ac.uk, mike.benton@bristol.ac.uk

*Corresponding author

WE are interested to see the new analyses and contributions by Vaux *et al.* (2018) concerning *Sphenodon*, and whether it can be called a ‘living fossil’ or not. We will focus on the part of their paper that is critical of our contribution on rhyngocephalian morphological evolution (Herrera-Flores *et al.* 2017) and comment only briefly on the remainder. We consider first our definition of ‘living fossil’, and whether such a definition can be meaningful or not, then we consider the specific criticisms offered by Vaux *et al.* (2018) of the analyses in Herrera-Flores *et al.* (2017).

DEFINITION OF ‘LIVING FOSSIL’

In their introductory remarks, Vaux *et al.* (2018) do not comment on our definition of ‘living fossils’ but dilate on the wider sins of other authors. The term ‘living fossil’ has had a long history, with many definitions and much debate about whether such a term is required or not. Vaux *et al.* (2018) do not like the term, and especially dispute that it can be applied to the New Zealand tuatara, *Sphenodon*. We agree with most of what they say. As we said in our paper (Herrera-Flores *et al.* 2017, p. 320), ‘The concept of ‘living fossils’ has been problematic since the term was coined by Charles Darwin (1859), as there is no such identifiable class of organisms’. Therefore, as we all agree that many definitions hitherto have been inadequate, a core purpose of our paper was to provide a computationally testable definition.

We repeat our definition here (Herrera-Flores *et al.* 2017, p. 320): ‘we propose a hypothesis that can be tested by computational morphometrics and phylogenetic

comparative methods (PCM): ‘a living fossil should show both statistically significantly slow rates of morphological evolution and it should be morphologically conservative.’” We went on to explain how these two features could be measured. We noted that morphological conservatism could be measured by both the distance in morphospace that a specimen or taxon lies from the centroid, or how different the specimen or taxon is from the average shape when using geometric morphometrics. Whether unusually slowly evolving taxa are called ‘living fossils’ or ‘unusually slowly evolving taxa’ is a matter of choice. The methods now exist to clarify the ‘unusually slowly evolving’ end of the spectrum of evolutionary rates in terms of statistically significantly slow rates.

In discussing whether *Sphenodon* is or is not a ‘living fossil’, Vaux *et al.* (2018) confirm that most previous authors have concurred that it is. Recent research using a quantitative metric unrelated to that in Herrera-Flores *et al.* (2017) has also identified *Sphenodon* as a ‘living fossil’ (Bennett *et al.* 2018). Vaux *et al.* point out that there is limited fossil evidence of New Zealand rhyngocephalians, and no evidence that the living species, *Sphenodon punctatus*, has had an unusually long duration. There is in fact no fossil evidence that it has had either a short or long duration, and this is probably not exactly relevant as, when using model-based approaches in macroevolution, rates of change are assessed from the sum total of available evidence. Our study was not concerned with *Sphenodon* alone, but with the wider clade Rhyngocephalia, comprising some 30 genera, and against which we assessed whether *Sphenodon* was close to the average morphology, or at the high or low end of calculated evolutionary rates.

CRITIQUE OF OUR ANALYSIS

Our paper presented two analyses: a study of evolutionary rates within Rhynchocephalia, and a geometric morphospace-disparity analysis. Vaux *et al.* (2018) did not comment on the first of these analyses, which is unfortunate as it addresses many of their concerns, but it is worth referring readers back to our paper, where they can see details of our analyses and results. In the evolutionary rates analyses, we used the data matrix of 74 discrete morphological characters from Apesteguía *et al.* (2014), and found, using PCM approaches, a scatter of significantly fast and slow rates of evolution across the tree; only two taxa showed consistently slow rates of evolution according to different analytical approaches, namely *Sphenodon* and *Priosphenodon*; for *Sphenodon* the slow-rate model was most consistent in all replicate analyses. Further, the consistently slow evolutionary rates shown by *Sphenodon* contrast with average and faster rates shown by the other four taxa within its subclade. The closest sister taxon of living *Sphenodon* is the Early Jurassic genus *Cynosphenodon*.

In commenting on the second analysis, the geometric morphometric study, Vaux *et al.* (2018) mis-characterize it in several ways. First, they stress the role of diet, and state that ‘The authors focused especially on the comparison of morphological variation with inferred diet.’ We did not. Our study was on morphology in an evolutionary sense. Vaux *et al.* (2018) further comment that the phylogeny in our figure 1 ‘indicated that tuatara are equally related to fossil genera that might have eaten plants and insects ... bivalves or crabs ... and seaweed’. These are interesting comments, but do not in any way affect our result, which was strictly morphospace-based, looking at landmark data on dentary shape.

Vaux *et al.* (2018) go on to give a lengthy, but inaccurate, description of our geometric morphometric morphospace analysis, implying that our purpose was to infer the diet of all the fossil forms with regard to modern *Sphenodon*. This was not what we did. In our plot, we simply labelled taxa by dietary categories previously assigned in the literature (we provided references). We then observed the distribution of the dietary groups in morphospace and related this back to the divergent morphotypes. We made no attempt to statistically link our analysis to diet, and we did not claim to do this; our plot is simply a visualization showing the distribution of dentary shapes and hypothesized diets according to categories established by other authors. Their misunderstanding of our paper is further clarified when they say, ‘Beyond diet estimated from dentary morphology, little consideration was given to other known differences among the rhynchocephalians sampled’. This is true: we made it clear throughout that our study was of disparity among

rhynchocephalians living and extinct based on mandible shape. The fact that mandible shape is related to diet is presented, but this is not the core of our argument. Nor did we claim, or attempt, to construct morphospaces that incorporated additional characters of the skull or skeleton.

Vaux *et al.* (2018) then provide discussions of the habitat occupied by modern *Sphenodon*, and the inferred habitats of the fossil forms. None of this discussion has direct relevance to our contribution. We did not claim to factor habitat into our analyses, and we made it clear that our analyses were focused on evolutionary rates in skeletal characters and geometric variation in dentary shape. Vaux *et al.* (2018) conclude this section by saying, ‘Therefore, focusing on a single trait is unlikely to provide a reliable assessment of evolutionary change in general ... or estimation of phenotypic stasis’. As readers of our paper will notice, we did not rely on a single trait. Our geometric morphometric analyses concerned dentary shape, and half the paper, which Vaux *et al.* (2018) ignore in their critique, analysed evolutionary rates in 74 skeletal characters that cover a broad range of cranial and postcranial anatomy.

We did not carry out landmark analysis on skulls or postcranial elements, and this would be an additional interesting study for someone to complete in the future. We had two reasons for focusing on the lower jaw for the landmark study: (1) such studies have been done frequently before by other authors on other vertebrate taxa (both fishes and tetrapods) and the studies have shown good morphometric discrimination between taxa; and (2) the mandible is most frequently preserved and so this maximizes the size of the data set; if we had added, say, skull, femur and humerus for landmark study, the data set of taxa would have been substantially reduced. As noted earlier, our first rates study used a data set of 74 characters that *did* include all aspects of skull and skeleton.

In two related comments, Vaux *et al.* (2018) critique our description of *Sphenodon*’s dentary as morphologically ‘average’ and conservative among fossil relatives. They suggest that we did not consider morphological variation expressed beyond principal components (PC) 1 and 2 (reflecting ~54% of overall shape variation). This is not true. In our analyses, we also calculated Procrustes distances, derived directly from the Procrustes aligned landmark data, to determine how close *Sphenodon* was to the average dentary shape for Rhynchocephalia. Procrustes distance (the sum of distances between corresponding landmarks from two shapes after superimposition) is the standard distance metric for shape (Zelditch *et al.* 2012) and is equivalent to utilizing information from all PC axes, not just the first two, or the first five, as suggested by Vaux *et al.* (2018). When we consider

these Procrustes distances, *Sphenodon* is recovered as the seventh most similar form to the average shape, out of 31 sampled taxa. We therefore suggested that *Sphenodon* does not have a divergent morphology and could be considered morphologically ‘average’ or conservative. We also wish to clarify that the dentary of *Sphenodon* is ranked as fourth closest to the centroid of PC1 (the primary axis of variation), not fifth as reported by Vaux *et al.* (2018).

Then, Vaux *et al.* (2018) make two important points that apply to every morphometric study: they urge the need for wide sampling to represent a decent average for the species, and they urge the need for precision and repeatability. In fact, we make these two points ourselves in our Method section, and we describe our strategies. First, Vaux *et al.* (2018) correctly note that there is at best a single mandible specimen for most of the fossil taxa, and so that was all we had. Using single specimens to represent the morphology of extinct species, and in some cases genera, is a common convention in palaeobiological disparity studies (e.g. Brusatte *et al.* 2012; Bhullar *et al.* 2012; Foth *et al.* 2012; Stubbs *et al.* 2013; Grossnickle & Polly 2013; Foth & Joyce 2016). Their first criticism then can only refer to *Sphenodon* for which hundreds of specimens exist in museums. We can answer their first criticism from our Method section (Herrera-Flores *et al.* 2017, p. 321): ‘[we use our own] pictures from 14 museum specimens of the extant *Sphenodon* to assess variation within this single taxon, and to determine where it falls in comparison with Mesozoic taxa. We performed a separate geometric morphometric analysis of all samples of *Sphenodon* to identify the specimen that best represents the average shape of its dentary.’ On the second point, we agree that precision is important, particularly in studies of variation within a single taxon like that presented by Vaux *et al.* (2018), where intraspecific variation will be the source of morphological variation. However, our study encompassed morphological variation across an entire order spanning millions of years. We argue that any minor intraspecific disparity, or variation incorporated by using published reconstructions and photographs of specimens in the literature will not introduce any large-scale bias. Care was taken when selecting what images to use, ensuring all specimens were appropriately orientated and not distorted. This is normal practice in studies such as these.

Vaux *et al.* (2018) go on to discuss the principles of fixed landmarks and semi-landmarks in geometric morphometrics, but we are not clear about their point. We are aware of the differences between fixed landmarks and semi-landmarks, and we have conducted many such studies before and followed standard protocol. It appears that Vaux *et al.* (2018) have misinterpreted our supplementary figure, and we did not fully explain the application of our semi-landmarks and curves. We used seven

separate semi-landmarks curves defining the outer margin of the lateral view of the dentary. These curves, and the number of semi-landmark points defining them, were consistently applied across samples. Each of these curves was anchored by the positions of fixed landmarks. During the generalized Procrustes analysis, the semi-landmarks on the curves were allowed to iteratively slide, minimizing the Procrustes distances between each specimen and the average shape (Gunz & Mitteroecker 2013).

Finally, Vaux *et al.* (2018) claim that the phylomorphospace presented in Herrera-Flores *et al.* (2017, fig. 3C) is erroneous. This claim is unfounded. Vaux *et al.* (2018) argue that the branching pattern within the phylomorphospace is a phenogram from dentary shape data, stating that ‘the phylogeny (more accurately a phenogram) was derived from the same dentary-shape-variation data used to estimate the principal components.’ This is incorrect. The tree topology superimposed in Herrera-Flores *et al.* (2017, fig. 3C) is not a phenogram derived from dentary shape data. As clearly stated in the original paper, figure 3C is a phylomorphospace showing the branching pattern of a phylogenetic tree in dentary shape morphospace. As explicitly described in the Method, this tree was derived from a maximum parsimony analysis using a cladistic character dataset with 74 discrete characters from the whole skeleton. We then projected this topology into the dentary shape morphospace using conventional methods (R package *phytools*; Revell 2012). As before, all our methods are widely used by numerous authors, and we followed established protocols carefully.

CONCLUSION

In summary, we confirm that the following criticisms of Herrera-Flores *et al.* (2017) by Vaux *et al.* (2018) represent misunderstandings or errors by the latter: (1) claims that we tried to infer diet from dentary shape data; (2) that we did not consider morphological variation beyond PC1 and PC2; (3) that semi-landmarks were applied inappropriately; (4) that the phylogeny used in the phylomorphospace was a phenogram from dentary shape data. Comments about using individual specimens, either photographs or reconstructions, to represent taxa in studies of disparity in the fossil record are interesting. However, the practice we applied is common in vertebrate palaeontology and the issue is not specific to our study or detrimental to the results presented in Herrera-Flores *et al.* (2017). We agree with Vaux *et al.* (2018) that care is required when selecting images for geometric morphometric studies.

The tuatara geometric morphometric analyses presented by Vaux *et al.* (2018) offer interesting results about

intraspecific variation in the modern and Holocene tuatara, but they present no criticisms nor results that refute the conclusions of Herrera-Flores *et al.* (2017).

Editor. Andrew Smith

REFERENCES

- APESTEGUÍA, S., GÓMEZ, R. O. and ROUGIER, G. W. 2014. The youngest South American rhynchocephalian, a survivor of the K/Pg extinction. *Proceedings of the Royal Society B*, **281**, 20140811.
- BENNETT, D. J., SUTTON, M. D. and TURVEY, S. T. 2018. Quantifying the living fossil concept. *Palaeontologia Electronica*, **21.1.14A**, 25 pp.
- BHULLAR, B. A. S., MARUGÁN-LOBÓN, J., RACIMO, F., BEVER, G. S., ROWE, T. B., NORELL, M. A. and ABZHANOV, A. 2012. Birds have paedomorphic dinosaur skulls. *Nature*, **487**, 223–226.
- BRUSATTE, S. L., SAKAMOTO, M., MONTANARI, S. and HARCOURT SMITH, W. E. H. 2012. The evolution of cranial form and function in theropod dinosaurs: insights from geometric morphometrics. *Journal of Evolutionary Biology*, **25**, 365–377.
- DARWIN, C. R. 1859. *On the origin of species by means of natural selection, or the preservation of favoured races in the struggle for life*. John Murray, London, 502 pp.
- FOTH, C. and JOYCE, W. G. 2016. Slow and steady: the evolution of cranial disparity in fossil and recent turtles. *Proceedings of the Royal Society B*, **283**, 20161881.
- BRUSATTE, S. L. and BUTLER, R. J. 2012. Do different disparity proxies converge on a common signal? Insights from the cranial morphometrics and evolutionary history of Pterosauria (Diapsida: Archosauria). *Journal of Evolutionary Biology*, **25**, 904–915.
- GROSSNICKLE, D. M. and POLLY, P. D. 2013. Mammal disparity decreases during the Cretaceous angiosperm radiation. *Proceedings of the Royal Society B*, **280**, 20132110.
- GUNZ, P. and MITTEROECKER, P. 2013. Semilandmarks: a method for quantifying curves and surfaces. *Hystrix*, **24**, 103–109.
- HERRERA-FLORES, J. A., STUBBS, T. L. and BENTON, M. J. 2017. Macroevolutionary patterns in the Rhynchocephalia: is the tuatara (*Sphenodon punctatus*) a living fossil? *Palaeontology*, **60**, 319–328.
- REVELL, L. J. 2012. phytools: an R package for phylogenetic comparative biology (and other things). *Methods in Ecology & Evolution*, **3**, 217–223.
- STUBBS, T. L., PIERCE, S. E., RAYFIELD, E. J. and ANDERSON, P. S. 2013. Morphological and biomechanical disparity of crocodile-line archosaurs following the end-Triassic extinction. *Proceedings of the Royal Society B*, **280**, 20131940.
- VAUX, F., MORGAN-RICHARDS, M., DALY, E. L. and TREWICK, S. A. 2018. Tuatara and a new morphometric dataset for Rhynchocephalia: Comments on Herrera-Flores *et al.* *Palaeontology*, published online October. <https://doi.org/10.1111/pala.12402>
- ZELDITCH, M. L., SWIDERSKI, D. L., SHEETS, H. D. and FINK, W. L. 2012. *Geometric morphometrics for biologists: a primer*, 2nd edn. Elsevier, 488 pp.