Reconstructed evolutionary patterns for crocodilie-line archosaurs demonstrate impact of failure to log-transform body size data

Roger B. J. Benson,1,2 Pedro Godoy,2,3 Mario Bronzati,4 Richard J. Butler5 & William Gearty6

MATTERS ARISING

https://doi.org/10.1038/s42003-020-01561-5 (2021)

Pseudosuchia includes crocodylians, plus all extinct species more closely related to them than to birds. They appeared around 250 million years ago and have a rich fossil history, with extinct diversity exceeding that of their living members1–3. Recently, Stockdale & Benton4 presented analyses of a new dataset of body size estimates spanning the entire evolutionary history of Pseudosuchia. They quantified patterns of average body size, body size disparity through time and rates of evolution along phylogenetic lineages. Their results suggest that pseudosuchians exhibited considerable variation in rates of body size evolution, for which they provided various group-specific explanations and asserted the importance of climatic drivers. This differs from two recent studies that analysed a substantial portion of pseudosuchian body size evolution and proposed that adaptation to aquatic life, a biological innovation of some subgroups, was the main driver of body size evolution regardless of data treatment, and one previous study did also not re

Measurements of real-world objects are taken in additive (or absolute) units such as millimetres (mm). This presents a well-understood statistical problem because biological variation is often multiplicative (=relative, or proportional), such that variance increases with increasing scale (heteroskedasticity; e.g. refs. 8,9). For example, a hypothetical group of rodents has species-mean body masses ranging from 100 to 200 grams, whereas hypothetical artiodactyls range from 100 to 200 kg. Both groups exhibit identical (two-fold) relative variation, but this is not reflected when additive units are analysed because the absolute difference between the largest and smallest artiodactyls is 1000 times that in rodents. Therefore, size variation among large-bodied species is substantially over-weighted unless measurements are log-transformed (see refs. 8,9). To ignore this, either by analysing or simulating trait data in a non-logged context, is to model a version of evolution in which it is as easy for a mouse population to evolve a body size increase of 10 kg as it is for an elephant population.

When using log units instead of additive units, identical proportional increases are indexed by identical numerical increases, entirely solving the problem (e.g. our rodents span from 2.0 to 2.3 log10-grams, cf. 5.0–5.3 log10-grams in artiodactyls). Therefore, evolutionary rate studies have routinely used log-transformed measurements for more than 70 years8–10. This is especially important when measurements span across orders of magnitude8, as with pseudosuchians which range from the largest species, Sarcosuchus imperator (skull length = 1650 mm) to the smallest, Knoetschkesuchus guimarotae (33 mm). This represents an estimated 125,000-fold variation in body mass, approximating that body mass scales with the cubic of linear dimensions (or 15,000-fold, conservatively reducing the skull of Sarcosuchus by 50% to account for its proportionally long snout). Although we illustrate the nature of this problem using extremes, it cannot be addressed just by excluding small- and large-bodied taxa from analyses, because variance increases as a continuous function of scale. We replicated the analyses of Stockdale & Benton4 using log10-transformed measurements. This resulted in a modified version of their PC1 size index that scales linearly with, and is strongly correlated to, log-transformed skull length (Fig. 1b; p < 0.001;
R^2 = 0.85; \( N = 202 \); Pearson’s product-moment). In contrast, the non-logged version\(^4\) shows a curved relationship (Fig. 1a) that upweights relative variation among large-bodied species—~80% of the variation in their index (y-axis) represents less than 25% of variation in relative (log-transformed) size. Due to space limitations, we focus only on macroevolutionary model comparisons and variation in evolutionary rates mapped to phylogeny. These analyses are central to the conclusions of Stockdale & Benton\(^4\) because they document variation in the tempo and mode of evolution among groups of different ages. Therefore, they provide process-based explanations that underpin the interpretation of their other analytical outputs, such as patterns of variation in disparity and average rates through time.

Re-analysis of rate variation among phylogenetic lineages\(^11\) using the non-logged PC1 size index returns similar results to those of Stockdale & Benton\(^4\) (Fig. 2a). Evolutionary rates returned by this analysis are measured in additive (or absolute) length units per million years and high rates occur in two contexts: (1) On the lineages leading to some large-bodied species such as the notosuchian *Razanandrongobe*, the phytosaur *Angistorhinus*, and the early crocodylomorph *Carnufex*; and (2) In some clades or grades of generally large-bodied species, including various Triassic pseudosuchians, Tethysuchia, Thalattosuchia and some eusuchians. However, these instances of high evolutionary rates are artefacts resulting from an underlying correlation of high rates with large body sizes (Fig. 1c; R^2 = 0.21; \( p < 0.001 \); \( N = 281 \) species; phylogenetic least squares regression (PGLS); R^2 = 0.46; \( p < 0.001 \); \( N = 559 \) species and nodes; ordinary least squares regression [OLS]). This is expected when data are not log-transformed because using non-logged measurements inflates the amount of evolutionary change inferred to have occurred among large-bodied species.

Analysis of our log-transformed version of the PC1 size index documents rate variation in relative length units per million years and is uncorrelated or very weakly correlated with variation in absolute size (Fig. 1d; R^2 = 0.03; \( p = 0.004 \); \( N = 281 \) species; PGLS; R^2 = 0.008; \( p = 0.19 \); \( N = 559 \); OLS). This rate variation shows a very different pattern to that of Stockdale & Benton\(^4\) (Fig. 2b), rejecting the occurrence of high evolutionary rates in large-bodied groups and species, including those listed above. This removes the need for hypotheses such as early evolutionary radiation of Triassic pseudosuchians, island endemism in *Razanandrongobe*, and viviparity in thalattosuchians\(^4\). These processes may well be important drivers of phenotypic evolution and species diversification in those taxa, but they did not result in above-background rates of body size evolution. Instead, we find a much more even distribution of high rates through time and among large- and small-bodied lineages, including high rates involved in the attainment of small body size in groups such as atoposaurids and shartegosuchids.

Nevertheless, biological interpretation of these patterns should be avoided because we find no support for the variable-rate model compared to a uniform-rate Brownian motion (BM) model (marginal likelihood = –477.2 compared to –470.7 for BM). Therefore, the variable rates model may be over-parametrised and should not be interpreted closely, overturning the
conclusions of Stockdale & Benton4, including their time series of rate variation, which is based on this model. We also find strong support for a constrained, Ornstein-Uhlenbeck (OU), model compared to BM (marginal likelihood ΔLOU = −463.5), consistent with previous studies3,6, but differing from Stockdale & Benton4. The constraint parameter of the OU model (α) is estimated as 0.016. This corresponds to a phylogenetic half life 12 of 43 million years, which is short compared to the study duration of ~250 million years. Pseudosuchian body size evolution is therefore highly distinguishable from Brownian (diffusive) evolution, consistent with the importance of functional and ecological limits to body size evolution at large phylogenetic scales.5,6 A speciational (kappa) model is best supported (marginal likelihood ΔLkappa = −454.0), but we disagree with the interpretation4 that this provides evidence of punctuational evolution, given that only a small proportion of species that ever lived are actually sampled in the fossil record.

Our analyses of log-transformed data reject key aspects of the conclusions of Stockdale & Benton4. More broadly, they demonstrate that the decision not to log-transform measurements introduces substantial errors to inferences of variation in the rate of evolution, and this should be accounted for in future studies.

Methods

We used the measurements provided by Stockdale & Benton4 to reproduce their published analyses of rate variation, also analysing a version in which the input data were log-transformed prior to analysis. Principal component analysis (PCA) with iterative missing data imputation was carried out in FAST version 3.1.7. We used the scores of the first principal component (PC1) as a body size index. Rate variation was evaluated on the time-scaled phylogeny of Stockdale & Benton4, using the ‘varRates’ model of BayesTraits version 3.0.213 and running our mcmc analyses for 2 million generations, with 50% of these discarded as burn-in (compared to 2 million generations with 10,000 discarded as burn-in by Stockdale & Benton4). Results were summarised using tools available at www.evolution.reading.ac.uk/VarRatesWebPP. We used mean scalar values as an estimate of evolutionary rates following Stockdale & Benton4, but our scripts allow the use of alternative rate summary metrics (Supplementary data).

Phylogenetic rate variation was visualised by painting rate colours to phylogenetic branches using functions from the R package ape version 5.0.314. We also used functions from ape to extract estimated body sizes at internal nodes of the phylogeny and to graphically compare body size to evolutionary rates (Fig. 2). We statistically tested the correlation of evolutionary rates to body size using ordinary least squares regression of node- and tip-rates to body size, and also using phylogenetic generalised least squares regression (PGLS) to compare summed root-to-tip rates to the PC1 size index, following e.g. ref. 15. This was implemented using custom code (Supplementary data) and the pgls function of the R package caper version 1.0.136.

Finally, to evaluate support for a variable rates model compared to alternatives, we conducted a model comparison analysis in BayesTraits11. This was based on stepping-stone sampling, running 10 stones each for 100,000 generations following 1,000,000 generations of burn-in, comparing varRates to speciational (kappa), Ornstein-Uhlenbeck and Brownian motion models.

Data availability

The datasets generated and analysed during the current study are available in the Figshare repository https://doi.org/10.6084/m9.figshare.15147306. This includes measurements from the supplementary data of Stockdale & Benton4, as well as other data required to replicate our analyses.

Code availability

The analytical scripts used in the current study are available in the Figshare repository17.

Received: 12 March 2021; Accepted: 25 January 2022; Published online: 25 February 2022

References

1. Bronzati, M., Montefeltrò, F. C. & Langer, M. C. Diversification events and the effects of mass extinctions on Crocodyliformes evolutionary history. R. Soc. Open Sci. 2, 140385 (2015).
2. Mannion, P. D. et al. Climate constrains the evolutionary history and biodiversity of crocodylians. Nat. Commun. 6, 8438 (2015).
3. Stubbs, T. L., Pierce, S. E., Rayfield, E. J. & Anderson, P. S. Morphological and biomechanical disparity of crocodile-line archosaurs following the end-Triassic extinction. Proc. R. Soc. Lond. Ser. B 280, 20131940 (2013).
4. Stockdale, M. T. & Benton, M. J. Environmental drivers of body size evolution in crocodile-line archosaurs. Proc. R. Soc. Lond. Ser. B 280, 20131940 (2013).
5. Godoy, P. L., Benson, R. B., Bronzati, M. & Butler, R. J. The multi-peak adaptive landscape of crocodylomorph body size evolution. BMC Evol. Biol. 19, 167 (2019).
6. Gearty, W. & Payne, J. L. Physiological constraints on body size distributions in Crocodyliformes. Evolution 74, 245–255 (2020).
7. Hammer, Ø., Harper, D. A. & Ryan, P. D. PAST: Paleontological statistics software package for education and data analysis. Paleontol. Electron. 4, 4–9 (2001).
8. Gingerich, P. D. Arithmetic or geometric normality of biological variation: an Empirical Test of Theory. J. Theor. Biol. 204, 201–221 (2000).
9. Hunt, G. & Carrano, M. T. Models and methods for analysing phenotypic evolution in lineages and clades. Paleontol. Soc. Pap. 16, 245–269 (2010).
10. Haldane, B. S. Suggestions as to quantitative measurement of rates of evolution. Evolution 3, 51–56 (1949).
11. Meade, A. & Pagel, M. BayesTraits V3.02: a computer package for analyses of trait evolution http://www.evolution.rdg.ac.uk/BayesTraits.html (2019).
12. Hansen, T. F. Stablizing selection and the comparative analysis of adaptation. Evolution 51, 1341–1351 (1997).
13. Paradis, E. & Schliep, K. ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. Bioinformatics 35, 526–528 (2019).
14. R Core Team. R: A language and environment for statistical computing (R Foundation for Statistical Computing, Vienna, Austria, 2020). https://www.R-project.org/.
15. Avaria-Llautureo, J., Hernandez, C. E., Rodriguez-Serrano, E. & Venditti, C. The decoupled nature of basal metabolic rate and body temperature in endotherm evolution. Nature 572, 651–654 (2019).
16. Orme, D. et al. caper: Comparative Analyses of Phylogenetics and Evolution in R. R package version 1.0.1. https://CRAN.R-project.org/package=caper (2018).
17. Benson R. B. J. et al. Reconstructed evolutionary patterns for crocodile-line archosaurs demonstrate impact of failure to log-transform body size. https://doi.org/10.6084/m9.figshare.15147306 (2021).

Acknowledgements
We thank Max Stockdale and Mike Benton for sharing data required to replicate their analyses, and for discussion. We thank Chris Venditti for writing with statistical advice on our preprint. We thank three referees for reviewing our manuscript.

Author contributions
R.B.J.B. designed the analyses, performed analysis and drafted the manuscript. P.G. performed analyses. R.B.J.B., P.G., M.B., R.J.B. and W.G. developed the concepts and contributed to the manuscript.

Competing interests
The authors declare no competing interests.

Additional information
Correspondence and requests for materials should be addressed to Roger B. J. Benson.

Peer review information Communications Biology thanks the anonymous reviewers for their contribution to the peer review of this work. Primary Handling Editor: Caitlin Karniski.

Reprints and permission information is available at http://www.nature.com/reprints

Publisher’s note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit http://creativecommons.org/licenses/by/4.0/.

© The Author(s) 2022