

Review of Body size prediction in scorpions: a phylogenetic comparative examination of linear measurements of individual body parts

I am pleased to provide a review of the contribution *Body size prediction in scorpions: a phylogenetic comparative examination of linear measurements of individual body parts*. I really enjoyed reading this paper, it is well written, clear, and makes a good justification of the utility of the research. It provides a useful supplementary dataset on scorpion size, and includes all the code required to replicate the analysis in a fully commented R script. I congratulate the author on a lovely piece of work. I note that I lack expertise in some of the statistics used in this paper, and thus have not felt confident in commenting on the veracity, or otherwise, of those, beyond the broad brush strokes below.

I have a few suggestions which the author may like to consider when resubmitting the paper. A number of these are listed below - of particular note may be the ones on the methods where I think, in places, individual choices could be more clearly justified. More broadly, the following sprang to mind when reading the paper, and could be worth of inclusion in the discussion:

- Given the justification of the utility of the methods used in the paper is that it is hard to measure total/carapace length, but that the models presented here are validated / compared to measured lengths from the literature, I wondered how much error there was likely to be in those empirical length measurements. Did all authors use the same measuring protocol? Is there any way to assess potential biases in that validation data? It may well be that the answer is no - and we have to accept this - but it could be worth adding a note to this effect!
- The models created here obviously work very well for buthids, and are presented as a more general tool for using proxies to assess total length. But what evidence is there that the patterns presented work outside this clade of interest? It could add value to the paper and enhance its wider adoption, if it were possible for the author to apply this to a small number of non-buthid taxa and demonstrate the efficacy of the approach in the non-target clade.
- Similarly, in several places the manuscript measures fossils - and this could be a very useful tool for fossils, I agree! But I would suggest that the author highlights that the older (and more phylogenetically distant) a fossil get, the less certainty we have that any of the relationships identified herein will apply to those taxa.
- In the discussion there was lots of comparison of different percentages of prediction power, but very little discussion of the absolute values. It might be just that I am a palaeontologist, but it struck me that the ranges of predictive power shown (i.e. ~10% accuracy) are actually really very good! Could it be worth highlighting that even if you were to choose the worst measure, in the most uncertain of cases, you're probably within 15% of the actual value?
- To maximise the impact of the publication, I suggest the author might like to make it more obvious where to find, and how to apply, the predictive equations derived herein. As written it could be that someone in a hurry missed that they are found in supplementary information file 2. If you would like people to use these more broadly, perhaps add a few sentences to the start of the discussion highlighting that this is where the equations are found, how to apply them, and what considerations the reader should keep in mind when they do so! It would also be possible, for example, to create an Excel spreadsheet or R script that allows the user to just enter a value and get the result from the model they want: adding something along these lines would further encourage the more widespread use of these findings.

None of these points are requirements for the paper to be published, although I do think they would allow the paper to reach and be used by the maximum number of people, so I would encourage the author to consider them. I hope they are useful. I am happy to be identified as a reviewer, and answer any questions arising: I do not write anonymous reviews.

-- Russell Garwood
russell.garwood@gmail.com

Abstract

-- technical problems persist for arthropods with complex body structures, as is the case for scorpions - is it possible to briefly be more specific about the nature of these problems? This would help justify better the improvement that is offered in this paper, and make it clearer why the chosen measurements introduced later do not have the same issues.

-- Otherwise, the abstract is clear, providing a concise overview of the methods, results, and then conclusion.

Introduction

-- The introduction provides a really nice overview of why these kind of measures matter, and the issues involved in collecting the required data.

-- In addition, for several non-flying arthropods ← several is generally used to mean a limited number (i.e. less than ten), whereas I think from context this may mean a larger number than that. Would many be more accurate?

-- typically expressed as the total length of the animal without appendages, which can be thus considered a homologous trait across the studied species ← this is only actually true if those animals have the same number of segments, right? Otherwise you are not comparing the same segment number. Is this true for the animals in question?

-- The same can be expected for flying arthropods - you can be more precise here and just say "flying insects", right?

-- size in various insect groups - can you say which insect groups without it taking up pages?

-- being also problematic for museum specimens without the considerable risk of causing damage to the individual, - is telescoping also an issue for museum specimens?

-- and some fossil records - I think this is a very generous reading of the state of the fossils! I would suggest that in the majority of fossils we would struggle to take measurements of carapace or body length with absolute certainty (of course, we also can't be sure that any relationships identified in Buthids would actually work for extinct taxa, but still, it is better than nothing when it comes to measuring size in long dead animals!)

-- Therefore, models that enable accurate prediction of body size measurements in scorpions [...] are valuable tools - I agree, and I think this section makes that case very strongly.

Material and methods

-- In general, the methods are detailed and very clear. I have a few points where I think clarification could be useful, which I list below.

-- and the posterior margins of the carapace (Fig. 1) - as shown on figure 1, the back of the carapace is a perfectly straight line, but I assume in living taxa there is sometimes a curvature on it. Is there any standard as to where this measurement should end in that case (the medial point on the posterior margin? The posteriormost point, wherever this is?), and do all of the publications you get measurements from follow the same protocol?

-- were obtained from scaled digital illustrations using the software ImageJ (Schneider et al. 2012) following the orientations of Sissom et al. (1990); this procedure was applied for the species *Aegaeobuthus cyprius* (Gantenbein & Kropf, 2000) and *Troglorhopalurus lacrau* (Lourenço & Pinto-da-Rocha, 1997) ← is it possible to repeat this procedure for one or more specimens with known measurements and assess the likely error?

-- Species means were - obviously it is not a mean if you have a single specimen. Reword to clarify this?

-- The tree was derived from nucleotide sequences of four DNA loci commonly ← how does this tree topology differ from those that are built using phylogenomic data, such as:

Santibáñez-López, C.E., Aharon, S., Ballesteros, J.A., Gainett, G., Baker, C.M., González-Santillán, E., Harvey, M.S., Hassan, M.K., Abu Almaaty, A.H., Aldeyari, S.M. and Monod, L., 2022. Phylogenomics of scorpions reveal

contemporaneous diversification of scorpion mammalian predators and mammal-active sodium channel toxins. *Systematic Biology*, 71(6), pp.1281-1289.

Obviously you use the tree you need to use (and I imagine studies such as above have incomplete species coverage) - but given the changeable nature of arachnid phylogenetics, it is worth highlighting if your phylogenetic results concord with, or differ from, those of others.

-- four loci were aligned independently using MAFFT v.7 (Katoh and Standley 2013) with default settings / trimAl (Capella-Gutiérrez et al. 2009) with default settings ← why the default settings? Are these the most appropriate for the task at hand?

-- This constraint aimed to ensure alignment with recent phylogenetic hypotheses for buthid scorpions (Ojanguren-Affilastro et al. 2017; Santibáñez-López et al. 2022; Štundlová et al. 2022) - but if you don't get that from your analysis, then could those other hypotheses be wrong? Why do you place more weight in those, than you do in your own analysis? There are many good answers to these questions, but I think you need to explicitly include them!

-- The resulting maximum likelihood tree was converted to ultrametric (branch lengths in million years) using the penalized likelihood method implemented in the *chronos* function from the *ape* R package (Paradis and Schliep 2019). To do so, calibration points were established based on fossil records, as outlined by Esposito and Prendini (2019). - I don't understand this! This tree is already, by definition, ultrametric if considered in terms of time, as all your terminals are living. So does this convert branch lengths from DNA into time? If so, how do you incorporate uncertainty in your fossil ages into your conversion? How does this align with molecular clock dates for the clades in question? I think just a few more details here would help me understand (I note that this is not my area of expertise, and so my comments may just reflect that, but the same may well be true of other readers of your paper!).

-- If this is the case, it would also be useful to highlight why a tree converted to time is most appropriate? Presumably branch lengths off DNA represent differences in sequence distance? Would we not expect long branches, along which lots of DNA evolution has happened, to similarly represent branches in which there has been more change in morphology? It may be that this is a standard for the field, and I don't expect the authors to deal with this disconnect as it is a big question, but this does leave me wondering if the results would differ if the tree used were the ML tree with branch lengths...

-- I note, your two calibrations are quite highly nested - what impact does this have on your dating? Is there any way you could get a date closer to the root? How does your root age compare with estimates based on molecular clocks?

-- with natural logarithmic transformation applied to all linear measurements - why?

-- Only one predictor was added at time - did you cycle through all possibilities of the second predictor?

Results

-- In general, I note I have limited experience of this kind of statistics, and my review will ultimately reflect that. I followed the results with ease, they are well structured and clearly written.

-- Figure 3 - is the point of this diagram how similar the values are? If so the chosen colour scheme highlights that well. But if you are trying to differentiate with this, what the best measures are, could this be clearer if the white part of the colour gradient was moved to nearer 0.8 to better differentiate the measures that matter?

-- However, the performance of bivariate models decreased slightly for species larger than 60 mm in total length, ← in males in particular (Fig. 5), there is one clear outlier of large size. Did you experiment with removing this to see if this individual had a disproportionate impact on the model fit? (I don't actually know if this is likely to be a concern - if not, please disregard).

Discussion

-- the use of carapace length as an indicator of body size is not always clear. ← what do you mean by "not clear"?

-- These findings indicate that while obtaining a reasonable indicator of size (length) in scorpions is relatively easy, the same cannot be said for reliably capturing information on overall shape. - can you say this from your data? If you were dealing with shape, you would do a PCA on shape data normalised by size, and it may be that in that PCA there is a clear pattern, but in your PCAs all of this detail is being lost because the impact of size overprints all of the other aspects so strongly.