Reviewer comments

I’ve reviewed the manuscript titled ‘Beyond CREA: evolutionary patterns of non-allometric shape variation and divergence in a highly allometric clade of murine rodents.’ This is an incredibly relevant and timely study given the recent work published on mammalian cranial allometry, and this paper offers a much necessary perspective. The authors present an enticing story: if CREA arises from biomechanical effects of cranial scaling, then deviations from CREA should have deviant biomechanical functions. This story, however, gets lost in methodological weeds. I think that this is a very important study that would be a significant contribution to the literature, but there are a number of methodological concerns that need to be addressed prior to publication. Specifically, a deeper discussion of the biology of the deviant taxa (Mastacomys, Hydromys, and Xeromys, etc) and how their biomechanical functions result in their straying from CREA, instead of the ‘will integration be reduced if we remove allometry’ discussion, is necessary.

Thank you very much for these positive comments and we were very happy to include the additional discussion suggested. This we have done this…XXX  
  
In the paragraph on lines 98-116, the authors state that it is a key question whether or not there are trends of cranial evolution away from the main allometric line, and then state that it is conceivable because allometry explains 36% of variation in their dataset. The authors have already answered this question however – about two-thirds of the cranial variation in the dataset is not related to size. The degree to which this non-allometric variation is related to function, I believe, is a very interesting and important question, one that the authors should focus on, but instead the authors ask how the removal of allometric variation will affect patterns of integration and modularity in their dataset. Once allometric variation is removed, the amount of integration will necessarily decrease because you’re removing a shape axis that describes covariation among landmarks – you’re removing covariation. This is the case for the removal of any axis of shape variation, there will be less covariation if you remove some of the covariation. This effect will be very large if allometry is aligned with one of the first PC axes (in your case PC1, line 169 & 282). On lines 324-327 the authors even state that the decreased integration in the residual dataset means that ‘size variation relates to an integrated response of modules across the whole cranium’, but to measure allometry they regressed size onto all of their shape variables, necessitating an ‘integrated’ response. Modularity will probably also increase because you’re removing among-module covariation, but the most-supported pattern of modularity might be different and this difference might be related to function (e.g., are the rostrum and molar landmark partitions found to be a single module when allometry is included, but found to be separate modules when allometry is removed? And is this related to the supposed stabilizing selection for gnawing?). The manuscript in its current form can’t accommodate this question though because there’s no analyses to find the most supported pattern of modularity either before or after the removal of allometry, only a comparison of the CR coefficient of the same modularity pattern in the allometry-included and -free datasets. These analyses could easily be substituted in, however. The authors on lines 98-100 state that allometry should be stronger in certain parts of the skull, which I agree with. Testing for allometry within each module, asking how much % variation it explains in different parts of the cranium, would be a highly relevant analysis for your questions.   
  
The authors seem to be unsure whether or not their removal of allometry is removing allometry, for example on line 172: ‘… to confirm that the allometry-free dataset lacks allometric information’. And on line 122: ‘if the allometry-free shape variation relates to the capacity of the cranium to diverge independently of allometry, we would expect…’ – what else would an allometry-free morphospace capture? If you’re unsure that the residuals of a linear model defining the shape-size axis is ‘allometry free’ then the rest of your analyses crumble because what, then, does the removal of allometry actually remove and therefore what can your subsequent analyses tell you about CREA? If the authors think that there’s additional allometry that’s not recovered in the linear model, then they should state that. The authors do state that there’s still ‘CREA-like shape variation’ in the allometry-free PCA, but this does not necessarily mean that your allometry removal didn’t remove all of the allometry, or that the initial allometry analysis didn’t capture all of the allometry, it just means that there’s additional face-length variation that’s not related to size. CREA states that proportional face length increases with size between species, not that all interspecific face-length variation is the result of size-related variation.   
  
On lines 111-112, the authors state, ‘high integration between modules (i.e., all modules co-varying strongly, also known as global integration’ and cite Bookstein 2015. Bookstein’s global integration analysis doesn’t consider covariation between landmarks, that was actually the motivation behind it, to measure ‘integration’ by leveraging the spatial structure of the data, not landmark covariances. The utilization of the global integration, specifically why this method to quantify integration is used over others (i.e., eigenvalue variances), isn’t discussed as much as necessary. This isn’t the most commonly used method, so it would be good to explain in more detail how it differs from other more conventional methods and why, given your questions, it’s the best choice. I would say that it makes sense to use it given that integration measured from landmark covariances will necessarily decrease once you remove allometry. Also, you should use species means for this analysis, not all of the specimens (line 237). Different sample sizes within species will lead to a miscalculation of the mean shape, whish is a crucial step to the global integration analysis.   
  
The authors state that they would expect to see Ornstein-Uhlenbeck type shape trends through time if diversification of the rodent crania is under functional allometric constraints (lines 119-120), but don’t perform any evolutionary model fitting. I agree that an OU analysis would be relevant (for the PC scores and size), so I was surprised that it wasn’t included in your analyses.   
  
Line 113-116: I’m confused with the wording of this sentence. What do the authors mean by ‘an underlying ability to change relative to the common allometry’? By this do you just mean morphological divergence in directions not aligned with the common allometric trajectory recovered by Marcy et al 2020? Or do you mean that allometry is evolutionarily labile?  
  
Lines 155-158: It would be interesting to see how the allometric axis differs when performing a phylogenetically-uninformed linear model. I say this because Mitchell, Sherratt, and Weisbecker 2023 astutely point out that when there’s a phylogenetic signal to body size (e.g., Cope’s rule), relevant size information will be removed when accounting for phylogenetic relatedness. Not necessarily in the main text, but it would be informative nonetheless.   
Lines 222-229: This is a peculiar way to measure phenotypic modularity. Why not use rPLS values from 2B-PLS analyses? It’s also unclear whether or not the authors re-ordinated the shape data per module before performing this analysis. Also, the panels D and E in figure 5 are the averaged rPLS value after each between-module 2B-PLS analysis, right? I don’t think this is mentioned in the text.   
  
Figure 3: Make aspect ratios 1 for the PCA plots.  
  
Line 486 and elsewhere: The authors never cite Cardini & Polly 2013, the original ‘CREA’ paper.   
  
  
Reviewer: 2  
  
Comments to the Author (see also attached annotated manuscript pdf)  
I revieved the manuscript entitled "Beyond CREA: evolutionary patterns of non-allometric shape variation and divergence in a highly allometric clade of murine rodents" that aims to compare patterns of allometric and allometry-free shape patterns of evolution in murine rodents.  
  
The present manuscript has (at least) a few aspects that positively impressed me like the presence of a Discussion section with almost no overstatements (except for the last paragraph) and the general aim to compare allometric and non-allometric shape variations (that is not so common topic in GMM and particularly interesting). Although in its current state the manuscipt is focused almost only on murine biology, these aspects clearly suggest that the manuscript has a good scientific potential.  
  
However, there a few aspects that worry me in this ms. First of all, there is a severe lack of literature references (I explicitly mentioned it for the Introduction in the attached .pdf, but also Materials & Methods are impacted as I indirectly highlighted in other comments). Following this line of reasoning, many procedures used in this research are not adequately motivated in the text (e.g., no references provided, no explanations for unusual procedures in GMM like removing PC1 in a morphological PCA to perform a matrix comparison, etc). Finally, the hypothesis of study concerning the presence of an evolutionary optimum not related to allometry appears to me to be described in a poor way and, more importantly, I suspect that the procedure adopted to (indirectly) validate such an hp is incorrect (and very dissimilar to the common application of multivariate phylogenetic comparative methods required to test for the occurrence of Ornstein–Uhlenbeck evolution and any other common evolutionary scenario that can be validated in shape data).  
  
I attach a .pdf version including all my comments in detail, hoping these suggestions/criticisms might help the authors to improve the manuscript and increase the scientific impact of this research.   
  
For all these reasons, I suggest a (severe) "Major revision" outcome for the present manuscript.