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.

**Thank you for this overall positive appraisal!**

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.

**This is a very reasonable concern and a real downside to working with a radiation with such a generalist diet. The fact that there are such few species that deviate from the overall allometric pattern makes it difficult to frame the deviations in a statistical framework. The suggestion to discuss the biology of deviant taxa is an excellent suggestion to deepen this discussion, but ultimately we are confined to post-hoc discussions rather than testable hypotheses.**   
  
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.

**I believe we were unclear here: is true that 64% of variation in the dataset is not explained by size, but this does not automatically mean that we should expect evolutionary patterns that lead to the divergence from the common allometric line. It is also possible that the species display shape variation that is unique to just them (e.g. through founder effects) and not part of any evolutionary pattern of co-variance, which is what this paragraph refers to. The sentence has now been clarified (p. X, line Y: “*This unexplained variation might be related to species-specific variation without any particular evolutionary patterning (such as vicariance or founder effects), but it also be attributable to evolutionary patterns not related to CREA”*)**

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.

**We agree that this is an interesting question, and have placed more emphasis on the question of how the different modules vary relative to each other (see below). A statistical approach to how the non-allometric variation is related to function is difficult beyond what we have already done – for example we predicted that we should see adaptations to hopping as a pattern emerging after allometric correction. This is because Australian rodents are such generalists that conventional analyses of function – e.g. a pgls analysis of shape ~ diet/locomotion – is not possible because such few species have these specialisations.**

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).

**This is correct and we were aware of this effect (both that it removes most of PC1 and will reduce integration). Our question was about the degree to which values for integration drop and values for modularity change when allometric effects are removed, so we see no issue here.**

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.

**If we interpret this comment correctly, the issue is that we were saying the same thing twice – allometry is measured as the as the co-variation of all shape variables with size, which automatically means that there is an integrated response of shape to size. Hopefully the following re-write on p. X, l. X-X addresses this unclear framing:**

*“This reflects the fact that allometric variation contains those patterns of integration that related to co-variation of all landmarks with size.”*

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.

**We have reported the degree of correlation between each pair of modules before and after allometric correction. The results were reported in Table 1 and discussed in the results section (under “Modularity and Integration”). We have now changed this analysis to report r-PLS values instead, but these results did not change. This does not answer the reviewer’s specific question of whether particular modules are retrieved as existing before/after size correction, but it provides a robust measure of how size correction affects cranial modularity. But we agree that our discussion missed the opportunity to discuss the functional implications of this. We have rectified this on P. X, l y:**

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.

**This is excellent advice and we have adopted it now in….**

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.

**It was indeed unnecessary to run these tests. These analyses were not because we suspected something to be wrong with your analyses, but because we wanted to make sure that everything lined up. This is something we do as a lab because it is good practice to be be paranoid and double-check everything, but it should not have gone into the published code or the manuscript and is now removed.**

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.

**Thank you for this useful advice! We have now extended the description of the method more broadly (p. X, lines Y-Z). The sample size figure was an error, the analyses were always done using the species mean and this is now rectified.**   
  
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.

**We have now added this analysis for PC scores and size.**

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?

**Apologies for the unclear wording – we have now adopted the wording suggested here (P. X, Lines Y-Z).**  
  
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.

**Good idea, thank you! Now done and presented in p. X, lines y-Z (or supps)**

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.   
**The vegan/mantel test pathway was just something we were used to from our earlier work. But 2B-PLS works well too, so we have replaced the analyses now. We have also fixed the caption to express that the averaged rPLS values are presented.**

From the phylo.integration notes: looks like we just do separate phylo.integration tests between all combinations

Note that phylo.integration performed on two matrices or arrays returns the same results as a phylogenetic variation of [two.b.pls](http://127.0.0.1:15731/help/library/geomorph/help/two.b.pls). It might be of interest with 3+ modules to perform separate phylogenetic integration tests between all pairwise comparisons of modules. This can be done, test by test, and the levels of integration can be compared with [compare.pls](http://127.0.0.1:15731/help/library/geomorph/help/compare.pls).

Figure 3: Make aspect ratios 1 for the PCA plots.

**Done – thank you for picking this up!**  
  
Line 486 and elsewhere: The authors never cite Cardini & Polly 2013, the original ‘CREA’ paper.

**Thank you for picking this up, this should not have been omitted. There was a lot of literature missing in this manuscript, which R2 also commented on. We have added a lot more references (see also comments on PDF).**  
  
  
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.

**Thank you very much for this positive appraisal!**

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).

**We have now added the literature recommended in the PDF comments and a few others.**

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).

**Thank you for this advice – we are commenting on these items below. [justify the phylo curve, probably add OU analysis]**

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.

**Reviewer comments in the PDF**

**Comment 1- Pg 5, line 47-50, sentence starting “The most widely discussed of these…”**

The description of this evolutionary pattern (i.e., CREA) in recent times was first included in Cardini and Polly 2013. However, several previous works pointed out such a relationship of braincase: facial scaling (see as an example some of the Radinsky's works I listed below concerning the mammalian order Carnivora).

Cardini, A. & Polly, P. D. Larger mammals have longer faces because of size-related constraints on skull form. Nat. Commun. 4, 2458 (2013).

Radinsky, L. B. 1981. Evolution of skull shape in carnivores. 1. Representative modern carnivores. Biological Journal of the Linnean Society 15:369-388.

Radinsky, L. B. 1982. Evolution of skull shape in carnivores. 3. The origin and early radiation of the modern carnivoran families. Paleobiology 8:177-195.

Radinsky, L. B. 1984. Basicrainial axis length vs skull length in analysis of carnivore skull shape. Biological Journal of the Linnean Society 22:31-41.

**Comment 2- Pg 5, line 54 “(in Press)”**

Now you can replace it and add the citation: I saw this paper a few days ago published on Biological Reviews! :)

**Comment 3- Pg 5, line 55-57 “closely related species tend to have similar diets and are thus…”**

N.B. This is not always true, just think about really famous case studies in evolutionary biology (e.g., cichlid fishes, Anolis lizards, Darwin's finches, etc) in which closely related species are very divergent in terms of dietary adaptations. Even without considering such extreme cases, resource partitioning as a way to reduce interspecific compatition is something that can be very common in nature (e.g., Stewart and Levin 1973; Pacala and Roughgarden 1982; Jenkins and Wright 1988, etc)

Partitioning of Resources and the Outcome of Interspecific Competition: A Model and Some General Considerations

Frank M. Stewart and Bruce R. Levin

The American Naturalist 1973 107:954, 171-198

Stephen Pacala, Jonathan Roughgarden, Resource Partitioning and Interspecific Competition in Two Two-Species Insular Anolis Lizard Communities. Science 217,444-446(1982).

Jenkins, K. J., and R. G. Wright. “Resource Partitioning and Competition Among Cervids in the Northern Rocky Mountains.” Journal of Applied Ecology 25, no. 1 (1988): 11–24. https://doi.org/10.2307/2403606.

DOI:10.1126/science.217.4558.444

**Comment 4- Pg 5, entire paragraph line 54-65**

N.B. Throughout this entire paragraph, you only used one/two citation(s) (that was/were not published when you submitted the manusctpt). Please add some references for the assumptions you are making. I am not saying that they are not reasonable, but that is a good standard for general readers.

**Comment 5- Pg 7, line 98-100 “shape variation independent of evolutionary allometry is expected…”**

N.B. For a non-expert of rodents as me, this part of the sentence does not seem so obvious. Can you please explain the reason behind it (maybe implementing some references)?

**Comment 6- Pg 8, line 118-126, entire paragraph starting with “If diversification of rodent crania is bounded…”**

N.B. I disagree with this paragraph. For example, you might be in presence of a pattern of shape variation dominated by allometry (that seems likely to be your case, even considering Marcy et al. 2020 results) that, after removing the allometric shape component, is still dominated by a (different) evolutionary optimum, produced by another (but non-allometric) functional constraint, which is blurred by the dominant impact of allometry on the morphospace/disparity through time plot. If such a weaker optimum is related to some aspects of ecological specialisation, then you will be in presence of a contraddiction with your sentence "By contrast, if the allometry-free shape variation relates to the capacity of the cranium to diverge independently of allometry, we would expect to see the highest morphological distances between species that have ecological specializations with divergent functional requirements, compared to those whose shapes follow the expected pattern of hyperallometric gracility." because you will be in presence of OU conditions of evolution in the allometry-free context produced by ecological specialisation.

Even in this paragraph, despite the presence of complex assumptions concerning the application of phylogenetic comparative methods and the interpretation of its biological meaning, no references at all are included in the text.

**Comment 7- Pg 8, line 130 “size from the dataset completely removes variation due to CREA”**

N.B. Sorry but I do not get it: can you please rephrase? I do not understand the practical (and also the theoretical) meaning of this part of the text.

Highlighted Pg 8, line 132 “and whether size-independent shape variation should show maximum shape disparity occurring”- no comment

**Comment 8- Pg 9, line 133-135 “between distantly-related species and/or species…”**

N.B. This comment is referred to the previous paragraph, this sentence, and the following parts of Mat&Met, Results and Discussion that are related to the same topic. I wrote it as the last comment at the end of my revision process.

For me, this hypothesis of study is extremely confusing and it is not described in a very clear way throughout the manuscript. In any case, if I understood well the authors are trying to validate the presence/absence of an evolutionary optimum in their shape (allometric and/or allom-free) data. To do this, they made various steps leading to tests performed on (morphological/phylogenetic) pairwise distances between species. At the end of the story, Figure 4 is graphically (i.e., qualitatively) evaluated to obtain some final inferences.

Now, the first issue with this procedure is that no references are provided in the manuscript in order to justify such a way to proceed (lines 245-258). I have never seen a similar framework used before and I think some explanations are required.

Perhaps even more importantly, the procedure suggested by the authors appears to be an indirect way to assess the presence of OU evolution without explicitely testing it. I do not see the point for using such an approach. To be more clear, I do not get the reason why the authors are not implementing in the manuscript model comparisons in order to assess the most likely model of evolution to describe their allometry or allom-free data. This is the standard way to proceed to validate the presence of an evolutionary optimum (i.e., OU evolution) or any other peculiar evolutionary condition, and it has been applied also in papers cited by the authors as Harmon et al. 2010 (but see for more recent theoretical discussions and examples on multivariate data papers like Adams and Collyer 2018; Clavel and Morlon 2020; Meloro and Tamagnini 2022; Bartoszek et al. 2022).

The authors need to implement explanations or alternative frameworks in order to validate this specific study hp, if they believe that is necessary for this manuscript, imo.

Dean C. Adams, Michael L. Collyer, Multivariate Phylogenetic Comparative Methods: Evaluations, Comparisons, and Recommendations, Systematic Biology, Volume 67, Issue 1, January 2018, Pages 14–31, https://doi.org/10.1093/sysbio/syx055

Julien Clavel, Hélène Morlon, Reliable Phylogenetic Regressions for Multivariate Comparative Data: Illustration with the MANOVA and Application to the Effect of Diet on Mandible Morphology in Phyllostomid Bats, Systematic Biology, Volume 69, Issue 5, September 2020, Pages 927–943, https://doi.org/10.1093/sysbio/syaa010

Carlo Meloro, Davide Tamagnini, Macroevolutionary ecomorphology of the Carnivora skull: adaptations and constraints in the extant species, Zoological Journal of the Linnean Society, Volume 196, Issue 3, November 2022, Pages 1054–1068, https://doi.org/10.1093/zoolinnean/zlab075

Krzysztof Bartoszek, Jesualdo Fuentes-González, Venelin Mitov, Jason Pienaar, Marcin Piwczyński, Radosław Puchałka, Krzysztof Spalik, Kjetil Lysne Voje, Model Selection Performance in Phylogenetic Comparative Methods Under Multivariate Ornstein–Uhlenbeck Models of Trait Evolution, Systematic Biology, Volume 72, Issue 2, March 2023, Pages 275–293, https://doi.org/10.1093/sysbio/syac079

**Comment to whole introduction**

N.B. I counted 13 cited papers in the entire Introduction section, with half of them that are mentioned in two sentences including lists describing previous papers on general topics (e.g., "This pattern, termed craniofacial evolutionary allometry (CREA), has been found in a diverse range of vertebrates representing 11 different orders, especially those of mammals (Bright, Marugán-Lobón, Cobb, & Rayfield, 2016; Cardini, 2019; Cardini et al., 2015)."; "This supports previous work suggesting that the strong allometry occurs as a result of stabilising selection on the rodent gnawing apparatus, which is highly specialized but allows for substantial dietary breadth (Druzinsky, 2015; Marcy et al., 2020; Zelditch & Swiderski, 2023).").

For me, there is a severe lack of references throughout the section: I think the authors should implement further literature, for instance, including different biological explainations that are suggested for CREA (see Cardini 2019 and references therein for some examples), clarifications concerning the assumptions they are making related to the application of specific PCMs (see my previous comment on the paragraph concerning Ornstein-Uhlenbeck), explainations concerning the importance of ecological specialisation and its interaction with allometry, etc..

**Comment 9- Pg 9, line 155-156**

Strikethrough ‘size less’ or

**Comment 10- Pg 11, line 180-187, sentence starting “However, variation characterised through ordination or allometric analysis…”**

N.B. Here, you are pointing out that these two visualisations are "different visualisations of allometry". I disagree since as you are saying one merely aims to show the differences between the largest and the smallest species (if such species are far from your allometric trajectory you will see very different shape changes compared to those that are really due to allometry) and the other is simply the shape variation pattern associated with PC1.

**Comment 11- Pg 13, line 243**

See my comment on this set of analyses + related Intro/Discussion parts.

**Comment 12- Pg 13, line 280-282, sentence starting “However, removing the full shape PC1 – and the…”**

N.B. How did you perform such a Mantel test??? The two matrices you are trying to compare have different dimensions (n-1 x k for full versus n x k for residual) violating the conditions to apply a Mantel test.

**Comment 13- Pg 16, line 283-287, sentence starting “Nevertheless, the relationship between the full shape and shape residual…”**

N.B. Forgetting for a second the previous comment on Mantel test, this seems a very empirical way to proceed and to make assumptions. Can a simple correlation test dispute evidences you are obtaining from a Mantel test (if we assume that such a test is correctly performed and perfomable)? I do not think so, but I would like to see examples from the literature of similar approaches used in research using geometric morphometrics in the text (maybe implementing them in the Materials and Methods section).

**Comment 14- Pg 17, line 308**

Strikethrough had, and replace with “had relatively”

**Comment 15- Pg 17, line 310-321 entire paragraph**

This is interesting!

**Comment 16- Pg 18, line 336-337, sentence starting “However, this could also be because…”**

Highlighted the sentence

**Comment 17- Pg 19, line 353**

See my comment on this set of analyses + related Intro/Discussion parts.

**Comment 18- Pg 22, line 425-427, sentence starting “For example, in the case of rodents, there are…”**

N.B. Concerning the role of gape angles on CREA, Tamagnini et al. 2023 recently demonstrated that extinct sabertoothed cats (possessing wider gape angles than their living counterparts) represent an exception to CREA, contrary to conical-toothed cats.

Tamagnini, D., Michaud, M., Meloro, C. et al. Conical and sabertoothed cats as an exception to craniofacial evolutionary allometry. Sci Rep 13, 13571 (2023). https://doi.org/10.1038/s41598-023-40677-6

**Comment 19- Pg 25, line 479-494, Paragraph starting “Counter to our expectations…”**

See my comment on this set of analyses + related Intro/Discussion parts.

Furthermore, contrary to the rest of the Discussion that seems pretty reasonable to me, this paragraph is full of potential overstatemens introduced by arbitrary evaluations that are unlikely supported by your results (e.g., "Instead, the asymptotic divergence pattern of shape reinforces the concept of rodent skulls evolving around an optimal shape unless a substantial change in either size or cranial function evolves."; "Overall, the pattern of “spikes” of divergence in otherwise limited morphospace (as demonstrated by a plateau of diversification after 4.2-5.7 million years) suggests an evolutionary pattern most consistent with a mean-shift Ornstein-Uhlenbeck process of limited diversification around a local optimum (Harmon et al., 2010).").

Perhaps, my criticism can be solved by implementing a solid explaination of the analytical procedure behind this part of the manuscipt and, in this specific case, of the link between Harmon et al. 2010 and the type of approach behind Figure 4 in the present ms.

**Comment 20- Pg 36. Not sure what figure, but figure title starting “Differences between the hypothetical shapes captured between PC1 extremes. A, plot of PC1…”**

Striked out “same as d” "‘allometry-free’ shape residual dataset"