**Brain-Climate reviews from *the Journal of Evolutionary Biology:***

Reviewer 1 provides incisive criticism of the approach used in the paper. In particular, the combination of using methods that assume stationarity over time with the aim to test for temporal trends in the data (i.e. deviations from stationarity) is a very serious concern. Reviewer 2, who identified herself as Susanne Shultz, has serious concerns about poor scholarship in relation to citation of earlier studies. In my own reading of the manuscript, I find that her and her co-authors' studies have been cited mainly as the source of data, but there is very little mention of how the present study relies on that earlier work for its aims and methods. For these reasons, I see no alternative but to reject the manuscript.

Thank you for submitting your study to JEB and I hope the reviews prove useful in revising your manuscript for publication elsewhere.

Sincerely,

Chris Klingenberg

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

Reviewer: 1

Comments to the Author

The manuscript asks whether brain size in mammals has changed over time, and if this can be explained by changes in temperature. Unfortunately, this means comparing multiple time series, which is a tricky subject because it is easy to find spurious effects. The authors use detrending to correct for non-stationarity, but this might also obscure the signal. I am thus not convinced by the results, and indeed suspect it would be close to impossible to disentangle the different possible drivers of brain size evolution with such data, but carefully building and interpreting a model would make this clearer. It may be helpful to try and get help from a tame econometrician.

1. Whilst it might be difficult or impossible to estimate body size, this does not mean it has no effect. This means that any results could just as well be because of adaptation in body size.

2. The ancestral state reconstruction is suspect. It assumes stationarity, i.e. there are no trends in size over time within a lineage. But you are testing precisely this hypothesis, so if it is true then you will not be able to see any effect. This is apparent from Fig. 5 where the two sets of data are different: this implies that the ancestral state reconstruction does not work, but a more formal test of this is necessary.

3. The authors confuse auto-correlation and stationarity. Detrending is a way of dealing with stationarity not autocorrelation. As the authors have multiple time series, multivariate time series methods (e.g. co-integration) may be appropriate, but I would be really careful as it is not clear what is signal or noise (e.g. if there is a trend in temperature and this drives brain size, then you need to have a method that will find this).  It might also help to note that differencing is equivalent to fitting Time as a covariate. If a trend in brain size is a result of a trend in mean temperature (say), then mean temperature should have a stronger explanatory power than Time. This will also mean data do not have to be averaged.

4. Does ocean/global temperature matter that much for individual species? Many can respond to climate shifts by migrating, rather than evolving, and global temperatures may not reflect the temperatures where a species lives. I know there's not a great deal that can be done (other than asking the NSF for money to build a time machine), but it might provide a good reason to not expect any effects.

5. CC is never defined in the text.

**Reviewer: 2**

Comments to the Author

The questions addressed in this paper: 1) does brain size increase across mammalian and hominin lineages, and 2) are these changes associated with climate change are of general interest to evolutionary biologists and especially evolutionary anthropologists. Therefore, there is much merit in the approach.

However, the current manuscript is offensive in the lack of credit given to the papers that provided the data and motivation for the analyses (Shultz and Dunbar 2012, Shultz et al 2012, Shultz and Maslin 2013). Firstly, this manuscript replicates the taxon by taxon approach to evaluate brain size increase (with similar results). Second, the introduction suggests that NO paper has incorporated autocorrelation in evaluation of hominid brain size increase and climate. This is precisely what Shultz et al 2012 did by detrending the data to residuals (and evaluating change in residuals brain size in adjacent time periods)- and found the SAME conclusion as this paper (albeit with slightly different methodologies). Thus, it appears that not only is credit not given to the analyses previously being done, but for the most part this paper REPLICATES these previous papers.

What is so absolutely infuriating about this manuscript is that if due credit were given to the papers that formed the analytical approach to the questions but that the improvements of this approach were clearly elucidated, I would be extremely supportive of an improvement to our previous papers. I am very pleased to see novel approaches for dealing with problematic paleo data. However, to present these analyses as completely novel is scientifically disengenuous. Therefore, I am not willing to engage my time and energy to figure out how these analyses contribute to the field. I feel it is the responsibility first of the authors to engage with the previous literature in a scholarly way and to be honest about what results replicate previous papers. If the authors are willing to do this, I would be very happy to reread the manuscript with an objective mind.

With best wishes,

Susanne Shultz

**Review of:**

Shultz, S., Nelson, E. & Dunbar, R. I. M. Hominin cognitive evolution: identifying patterns and processes in the fossil and archaeological record. *Philos. Trans. R. Soc. B Biol. Sci.* **367,** 2130–2140 (2012).

While I was at first surprised I did not find this study in all my literature searches, I think this may be because it is listed as a review with a title that did not suggest a study of brain-climate interactions. Despite not having read it until after the JEB rejection, its quite remarkable how similar our approaches are in many ways.

**Their procedure:**

*CC vs Time:*

* Linear regression of CC vs. time (to test for brain size change over time and to get CC residuals (NOT relative to body size))
* Group time periods into arbitrary 100-200ky blocks (with abs CC mean or resid CC mean calculated for each block)
* Do a whole bunch of pair-wise tests between blocks without correcting for multiple comparisons.
* Repeat for different subdivisions (super-species of hominins), again without correcting for multiple comparisons.

*CC vs Climate:*

* Linear regression? (gradual change test)
* Step/pulse tests: calculate mean and SD of each climate measure for 100ky prior to specimen date. Run correlations between CC (abs and resid) and climatic measures (each separately, again without correcting for multiple comparisons)
  + Also did step-wise regression with all climatic measures (but results not reported bc none were sig)

**Main differences between our Brain-Climate manuscript and Shultz et al 2012:**

1) \*\*\*Only hominins included. No comparative sample of other mammals.

2) Used Miller foraminfora dataset (sealevel), not Zachos (temp). Diff interpretations of similar measure?  
  
3) Detrended data by taking residuals of CC against time. However did NOT use 1st-diff (and thus did not need to detrend climate data). The fact that we used 1st also necessitate averaging for each time point (e.g. @1.8Mya).  
Also, their hominins only go as far back as 3.2My, ours extends to *Ar. ramidus* @ 4.4Mya and includes robust australopiths.  
Also we only pooled all hominins together (didn't group by "super-species", which seem rather arbitrary).

4) Intervals are 100ky OR 200ky WITHIN the same time series! (only first two are 100ky). Also these are arbitrarily set bins going back from modern day (0ky). My intervals are 200ky, 400ky, 1my going back from the points at which the fossils are dated.  
Also, only using small bins means many bins have only several data points.  
  
5) Climate measures are mean & SD sea level, and mean & SD aeolian dust. They are all independently tested without correcting for multiple comparisons. I put all measures in a single model that corrected for multiple factors.  
Also, they don't include slope (Rate).  
Also, they use correlations to test brain-climate relationships, instead of linear models.  
  
6) While they did technically detrend their CC data via regression, they did not make this an explicit point of their paper as to why this is critical. We go into much more detail and explain why, which is valuable to the field.

**What to do next:**

* **1) Make minor revisions to the manuscript.**
* Emphasize the novelty of our diverse comparative approach for testing brain-climate interactions in the context of human brain evolution.
* Give due credit to previous Shultz publications. Explicitly discuss the contributions of Shultz et al 2012 in the Intro (time-series section) as well as how our results compare to theirs in the Discussion (largely validating the lack of hominin brain-climate interactions (even more so that Shultz’s inconsistent results), but with novel non-hominin brain-climate interactions).
* More clearly (but briefly) discuss the points brought up by Reviewer 1 throughout the manuscript. Particularly, state the assumptions of our ancestral state reconstruction model (which uses Brownian motion) and how this may have affected the results (cite Pagel chapter).
* Recheck wording regarding the topics of trended data, temporal autocorrelation, and detrending procedures.
* **2) Send revised manuscript to Shultz to request her feedback and offer an olive branch of peace.**
* Brian will send out a draft email to all co-authors before sending it to Shultz.
* Include her in the acknowledgments if she accepts.2
* **3) Send out for review at JHE (A.S.A.P.).**
* JHE is the right length and will reach the desired biological anthropology crowd. Will also have a better chance of being edited by someone more familiar with these topics.
* If submitted soon, might be published in time to generate discussions at the AAPA 2017.