

Leibniz-Institut für Evolutions-
und Biodiversitätsforschung

Dr. Eli Amson
DFG postdoctoral researcher

Tel +49 30 889140 8339
Mobil +49 152 3477 5360
Mail Eli.Amson@mfn.berlin

www.museumfuernaturkunde.berlin

Reply to Recommender

Berlin, 4 September 2018

Dear Recommender,

We believe that we have addressed all the comments of both Reviewers and of the additional PCI Recommender. This involved re-running some of the analyses and clarifying and expanding most sections of the manuscript. We trust that the revised manuscript (RM) is greatly improved. Please find an accompanying version of the Word document with tracked changes (which was formatted using the PCI preprint template), and below our detailed response. The revised manuscript was also posted on Biorxiv ([biorxiv.org/cgi/content/short/318121v2](https://www.biorxiv.org/cgi/content/short/318121v2)).

On behalf of both co-authors,
Eli Amson

Reviewer 1: anonymous

The manuscript of Amson and Nyakatura is an interesting contribution to the existing literature on the relationship between trabecular and cortical bone and behaviour of extant xenarthrans, and tests whether the behaviour of fossil taxa can be reconstructed. In their previous publications, the authors have been the first researchers to apply modern methods to investigate trabecular bone structure in non-primate taxa, and I am pleased to see the potential utility of this aspect of bone structure explored. The humerus and radius are analysed amongst extant taxa which differ in both their locomotor mode, and in the frequency and mode of fossorial behaviour. Several fossil xenarthrans were analysed in order to determine their locomotor and fossorial behaviour. Although, locomotor and fossorial behaviour could not be not confidently assigned to the fossil taxa included in the analysis, Hapalops was found to have sloth-like morphology. I am pleased to see the authors discuss the complexities of this methodology, including the problems arising when interpreting bone structure in such large-bodied, extinct species. I have divided my comments on the manuscript into major and minor suggestions, and have also noted spelling and grammatical errors.

Major comments

1. The manuscript would greatly benefit from a more detailed and nuanced interpretation of the various bone parameters measured. Cortical and trabecular bone are analysed collectively in the manuscript, however, the two regions of bone perform very different functions. Cortical bone cross sectional geometric properties reveal the potential for bones to resist compressive and bending forces, whereas trabecular bone structure is adapted to compressive forces at the articular surface. As such, the two types of bone would not necessarily be expected to co-vary directly, as assumed by the authors. The manuscript should more clearly explain how these two different bone structures reflect mechanical loading.

It was not our intention to convey that trabecular and cortical parameters necessarily co-vary. We agree with the reviewer on this regard, but we do not believe that this is incompatible with the inclusion of parameters from both compartments in a single discriminant analysis. But one can argue, however, that the potential absence of covariation between the trabecular and diaphyseal parameters is an argument to keep all parameters separate (i.e., not using the mean bone fraction for instance). This is what we did in our initial univariate and multivariate approaches. The introduction was revised to clarify the matter.

2. A separate, but related comment is that the biomechanical hypotheses tested are not clearly explained. I understand that often the kinetics/kinematics of behaviours such as climbing and digging are unknown, but the authors should

elaborate on the differences that would be expected for their behavioural categories. For example, what differences might there be in the variability of loading (as measured by DA) between fossorial and non-fossorial groups? Or, how might the overall the magnitude or orientation of loading differ between behavioural categories? Although not feasible to incorporate in the present study, I wonder if elements of the hindlimb might be informative for future studies. The two behaviours of interest in these groups are arboreal/terrestrial locomotion and fossorial behaviour. I assume that digging in these groups only, or at least primarily, involves the forelimb, in which case you would expect the hindlimb to be influenced only by locomotion. As it stands, the elements in this study are influenced by both locomotion and digging, thus the morphology is likely due to a combination of two different loading regimes.

The introduction was edited to state the expected main biomechanical differences among the functional categories and arguably associated differences in the structural parameters. Digging is indeed practiced mostly with the forelimb in xenarthrans and we fully agree that examining the hindlimb properties should be relevant to understand differences between terrestrial and arboreal lifestyles. What explains our focus on the forelimb (since the first analysis focusing on extant taxa, Amson et al. (2017a)) is that we were also interested in comparing the different degrees of fossoriality among armadillos.

3. The cross-sectional geometry of the elements included in this study are highly complex, and at the sites analysed there are large muscle attachment sites. I would suggest that the authors discuss the possible impact of these muscle attachments sites on the cross-sectional geometry results, and for future analyses consider sampling locations on the diaphysis without prominent entheses.

A paragraph was added to the discussion to suggest this approach for potential future analyses.

4. Information should be included on sample size and extant species in the sample, and the details of CT scanning of the extant sample (i.e. where they were scanned and at what resolution).

We revised the Material and Methods section to include this information (resolution was already available in the initially submitted SOM 1, but we admit that it was not properly advertised). Because the information regarding extant species' sample size and CT scanning details was given in Amson et al. (2017a), we added a sentence to the RM to invite the reader to refer to it.

5. I would like to see more information about the VOI placement protocol. The methods say that ROIs were selected from the centre of the epiphyses, however,

looking at the images in Amson et al. 2017, this doesn't appear to be the case for either the humeral head or the radial trochlea, the MC3 was not included in this previous publication. A clearer description, and preferably a figure, should be included to explain further the VOI placement protocol.

We expanded the relevant section of the RM. It is unclear to us why the Reviewer considers that the humeral head and radial trochlea ROIs do not correspond to the centre of the respective epiphysis. The Reviewer refers to "looking at images". If she/he only refers to the Figure 3 Amson et al. (2017a), we invite her/him to also look at the 3D pdfs given as SOM of Amson et al. (2017a). We did not acquire trabecular data for the Mc III (so no ROI placement was made in its case, only a mid-diaphyseal section, as explained in the manuscript).

6. p10, line 223: Was the total volume used here the size of the VOI? This is unlikely to be a good size proxy, because the VOIs were not scaled to the size of the epiphysis, rather as large a VOI as possible, avoiding cortical bone, was placed in the epiphysis. Although the TV is not used as a size proxy in the analysis, a measure of the size of the epiphysis would be more appropriate than the size of the VOI.

As explained in the Material and Methods of the initial manuscript, indeed TV is used as one of the possible size proxies. It is not precisely clear to us what the Reviewer refers to with "a measure of the size of the epiphysis." We will assume that by this it is referred to a length such as the femoral head height, as in Ryan & Shaw (2013; Proc R Soc B 280: 20130172) for instance. Given the quite disparate morphology of the epiphyses among xenarthrans (e.g. for the humeral head, very spherical in extant sloths, much less so in armadillos), we did not expect for such a measurement to necessarily be a better size proxy. But we will consider it for future analyses. In any case, as the Reviewer points out, TV was eventually not used as a size proxy in the present study.

7. p12, line 265: Were the parameters normally distributed after log-transformation?

Yes. The corresponding sentence was edited in the RM to clarify it.

8. The paper should include a results table with the mean values for each taxonomic group or species, and the results for each fossil.

Two additional tables (revised Tables 2, 3) were included to the RM.

9. p13, Univariate Comparisons: The focus of this section is on the fossil taxa, but it would help the reader a brief description of how the extant groups differ from one another was included, for all parameters discussed.

This section was revised accordingly.

10. Not enough information is given concerning the discriminant function analysis – I would expect the paper to include a table reporting data from this analysis, and additional information in the text. For example, a table with predicted group membership should be included for both the training data and the fossil taxa. It would also be informative to include the contribution of the variables to each function to better understand which variables are driving between group differences. How were the extant taxa grouped – it is unclear whether this is at the species level, generic level, or by a behavioural classification?

A table (revised Table 4) with predicted group membership for fossil taxa was added to the RM. The training data is now reported in an additional SOM (revised SOM 4). Percent of explained between-group variance of each axis was added to the revised Figure 5. Contribution of the variables to each function is now provided in the revised SOM 5, and summarized in the text. The extant taxa are grouped by lifestyle classification (which, as explained in the text, matches main clades except for sloths). A sentence was added to the revised Results to clarify it.

11. What is the potential influence of correlation between variables on the DFA, and on the PCA used for *Hapalops*? The included trabecular parameters are likely to be correlated with one another, for example BV/TV and Tb.Th.

We revised our whole procedure to include a consideration of the correlation among the variables included in each discriminant analysis. As predicted by the Reviewer, several trabecular parameters were highly correlated. We accordingly amended the Material and Methods section as well as the Results (and correspondingly the Fig. 5). The new results are very similar to the initial ones. One of the modified analyses was that of *Hapalops*. The latter does not require the dimension reduction we formerly applied using a PCA anymore.

Minor comments Abstract: The authors overstate the sensitivity of trabecular architecture by using the phrase “extreme accuracy and sensitivity” in the abstract; in p1 line 44 “great accuracy and sensitivity”; and in p1 line 47 “great plasticity”. Studies in primates have had very mixed results, in many cases the relationship between trabecular structure and behaviour is unclear. I would recommend these phrases are adjusted to reflect that it is not known how accurately trabecular structure reflects loading.

We agree that the analyses examining bone structure and primate lifestyles yielded mixed results, and revised our Introduction to express it. However, the sentences the Reviewer refers to (Abstract and beginning of Introduction) actually deal with the more general concept of plasticity of

trabecular and diaphyseal structure. In contrast to the relationship to lifestyle across various species, we consider that the plasticity of bone structure is well demonstrated experimentally, and therefore prefer to keep the sentences in question unchanged.

P4, line 65: Tsegai et al. (2017) used the cortical thickness method developed by Treece et al. (2010; 2012). It is important to note that the focus was cortical bone thickness at the articular surface, rather than diaphyseal structure.

The sentence was revised accordingly

p4, line 88 and p7, line 153. Anteaters are described here as intermediate, it should be more specific, is this intermediate in their fossorialism or terrestrial/arborealism?

We considered anteaters as intermediate both in their arboreality and fossoriality. This was clarified in the RM.

p9, line 209: Change “trabecular” to “cortical”

Corrected in the RM.

p17, line 401: Include other publications from the primate literature, as there are many studies which find DA, or primary trabecular orientation, to be informative (e.g. Ryan and Ketcham, 2002; Griffin et al., 2010; Barak et al., 2013; Su et al., 2013).

Added (along with additional publications to the RM).

p19, line 460: This is an important point, which could be expanded upon. Is there any evidence for this in extant xenarthrans?

There is little evidence due to the lack of experimental data. We added some speculation (identifying it as such) in the RM.

Spelling/grammatical comments:

All suggestions were included to the RM.

Reviewer 2: Andrew Pitsillides

Overall comments: This work in this manuscript seeks to establish relationships between known lifestyles and cortical and trabecular parameters, in a range of bones of extant animals from a specific clade. The intention seems to be, to use any relationship that emerges to make predictions regarding the lifestyles in extinct clade members, based solely upon scans of their fossilised bone remnants. The authors present a combined analysis of cortical and trabecular parameters of the

xenarthran forelimb, based on data from one of their group's previous publications (trabecular bone of extant xenarthrans) supplemented with novel trabecular data from extinct tardigrades (ground sloths) and cortical data from all xenarthrans. Whilst the previous study, predictably found that forelimb function matched the phylogeny – thus, that “all armadillos dig, all sloths climb and all anteaters do a bit of both” – this new manuscript, in contrast, finds instead that ground sloths don't match the lifestyle of the extant members of their clade.

Major comments 1. The authors claim to have identified the ‘challenges’ to making any such predictions with security. These are stated to be: i) the imperfect lifestyle discrimination for the extant animals, ii) the difficulties of scaling of these parameters for extinct animals, which are outside the range for extant animals and iii) the classification of the values from extinct animals as outliers with respect to the extant animals. This reviewer is not expert enough in this area to know whether these ‘challenges’ could have been predicted before undertaking these analysis, nor whether their identification alone is novel enough to warrant reporting. Both need to be addressed by the authors and then an expert palaeontologist should verify the robustness of their response. This applies particularly to their conclusions (and throughout); for example (line 120-123), where it was unclear to me what the purpose of this study was, unless it was to unambiguously identify the reasons that the reconstruction [lifestyle] of the extinct Xenarthrans cannot be achieved by the means they have used?

The chief purpose of the study was not to identify challenges, but to try to reconstruct the lifestyle of several extinct sloths. The end of the Introduction was revised to clarify it. We just consider that it was relevant, especially in an analysis dealing with a fairly new approach, to discuss the challenges it involved. Furthermore, lifestyle reconstructions were not entirely successful, which also motivates discussing the limitations of the approach and the problems we encountered/identified. It was not our intention to claim to have identified the ‘challenges’, but some challenges. In the initial MS, for instance at the beginning of the Discussion, we therefore referred to “at least three obvious causes”. None of the three identified ‘challenges’ were entirely predictable before undertaking this analysis: i) most diaphyseal parameters were never acquired for xenarthrans, their potential for lifestyle discrimination was hence hitherto unknown; ii) similarly, scaling for these parameters among xenarthrans was not known, and studies describing allometry in trabecular parameters (Doube et al. 2011) did not include “ground sloths” either; iii) finding outlying parameter values for some of the extinct taxa (e.g., extremely compact diaphyseal structure in the putatively semi-arboreal *Hapalops*) was completely unexpected.

2. The authors state (line 43) that trabecular bone reacts to loading very sensitively, which indeed it does rather rapidly too, to provide a 'functional signal'. Does this bring into question the use of its analysis as a readout of likely 'lifestyle' in extinct animals? These extinct animals would have to have died in perfect 'health' in order that direct comparison with extant, presumably healthy, animals is to be made. This might indeed be the case and so the likely reason for their extinction is rather important and should be stated if known. This point also raises another major worry; might this statement be deeply misleading. I would recommend the authors make it clear, that, although the fact that bone is typically responsive to loads, which parts of the bone respond to what extent is still very much unclear. As the review of trabecular bone functional adaptation cited by the authors (Kivell, 2016) states: "bone functional adaptation is not sufficient (but is all we have)". I have no problems with the authors trying to infer lifestyle from the information they have, but they should make it clear that this approach is (necessarily) limited.

We revised the Introduction to temper the references related to a relationship between bone structure and lifestyle (see answer to one of Reviewer 1's comments above), to make clear that this approach is limited. While one might expect some individual variation to affect bone structure to some extent, one might also expect some broader differences to be present between taxa of wholly different lifestyles. We added a sentence to the RM's Introduction to state it. We are not sure to understand what the Reviewer refers to with "reason for their extinction". We certainly do not assume that, because they belong to an extinct clade, the sample individuals might have been unhealthy. That is why we do not think it is relevant to discuss the cause of extinction of the sampled taxa. We did not observe any bone diseases in the specimens. We added a sentence to the RM (Material and Methods) to state it.

Minor comments 1. Abstract: The last sentence is not clear and should be re-written. I think an issue with the phrasing is due to the use of the word 'challenges'.

The sentence was rephrased in the RM.

Introduction: 2. I would like the authors to make the aim of the paper clear in the last paragraph of the introduction. Are they trying to reconstruct the lifestyle of extinct xenarthrans based on a combination of cortical and trabecular parameters (as I was led to believe until line 111), are they presenting a new method to reconstruct function from trabecular and cortical parameters (as suggested by the review of previous similar methods in the paragraph starting on line 55) or are they identifying challenges related to reconstructing lifestyle from trabecular and cortical parameters (as they state at the end of the introduction)? This issue occurs elsewhere and my preference would be for this to be revisited (throughout) to add greater meaning to the conclusions. Current phrasing suggests that the authors

have set out to identify whether there are 'challenges' rather than address questions that could realistically have been addressed using the samples available.

A sentence to clearly state the goal of the analysis (i.e., reconstructing lifestyle of extinct sloths) was added at the beginning of the last paragraph of the revised Introduction (see also answer to Major comment 1 above). The Conclusion was also revised to clarify the matter. The Discussion already started with a sentence related to it, so we consider that it was already clear in the initial manuscript.

3. Line 12: "relative number" relative to what?

The word was deleted in the RM.

4. Line 45. accepted for 'excepted'

Corrected in the RM.

5. Line 66: "same" it is not clear to me whether this refers to the current study or to Gross et al

Clarified in the RM.

6. Line 70: "distinct zones of different arrangement" I presume this refers to e.g. what is sometimes called the vertical and horizontal trabecular columns in the femoral neck (Hammer, 2010, Annals of Anatomy, "The structure of the femoral neck: A physical dissection with emphasis on the internal trabecular system") but this is not clear... could the authors please clarify?

Indeed, this is what was meant. It is clarified in the RM.

7. Line 79: I do not understand why a medulla full of trabecular bone helps withstand compression. Could the authors elaborate?

This was deleted in the RM.

8. Line 85: Could the authors please expand on the findings of their previous paper (Amson et al, 2017a) here? I would help set the context for the next paragraph.

This was done in the RM, but in the next paragraph, in accordance with a comment from Reviewer 1 (see above).

Page 5, line 89-92: it is not clear to this reader exactly what point is being made here. Suggest rephrasing.

This was rephrased in the RM.

9. Line 104: I presume the phrases/terms autapomorphic nature, phylogenetic signal and ecophenotypic character are used appropriately? How can

“ecophenotypic character” be a “rationale”? And what is an ecophenotypic character? Please clarify.

Clarified in the RM.

10. Line 91: I understand what the authors mean by “likely not true anymore”, but this should be clarified, e.g. by adding a clause about why this is “not true anymore” at the end of this sentence.

Clarified in the RM.

11. Line 119: It seems to me that the large-sized ground sloths exceed extant xenarthrans by more than an order of magnitude in body mass?

Corrected in the RM.

Materials and methods 12. The exclusion criteria on page 8 are not readily understood. In general, the methods are very discursive. I am not sure whether this is acceptable or not. Might it be more meaningful to move some of this apparent ‘validation/deliberation’ (pages 10-11) to the discussion?

Page 8: exclusion criteria were clarified. P 10-11: because of a change in the analysis related to one of Reviewer 1’s comments, the ‘validation/deliberation’ was not relevant anymore, and hence excluded from the RM.

13. It would be good practice if the authors provided their R-code online as supplementary material, it may be useful for others to replicate their results/use their method on a different data set.

The R script devised for this study merely consists of data loading and running of tests using published functions. Each R function is accompanied by a user guide, and we do not believe that our script will be very useful for other users in that regard. We did, though, wrote a few simple functions for the image analysis software Fiji, and provided the code in the initially submission.

14. Line 146: I believe this is a typo: “scanning resolution ranged from 0.03 to 0.123 micrometer” - shouldn’t this be 0.123 millimeter or 123 micrometer? To my knowledge, typical microCT scans performed in our lab on mouse tibiae result in a pixel size of several micrometers to several tens of micrometers – 0.03 micrometer=30 nanometer which I don’t think is realistically possible in a reasonable amount of scanning time (especially for large species such as ground sloths) or necessary to gather trabecular geometry information?

Corrected in the RM.

15. Line 203: I believe this is a typo: did the authors forget to abbreviate global compactness to GC in this sentence?

Corrected in the RM.

Results 16. Could the results be shortened?

It is unclear to us why they should (they occupy less than 2 formatted pages, without the figures)

17. Univariate comparisons: It is unclear to me how the authors choose the anatomical locations (% length) to compare the cortical parameters in humerus and radius to each other: sometimes it's mid-diaphysis, sometimes 72% and sometimes 35%. I assume Figure 3 shows mid-diaphyseal data? This should be made clearer. If the authors compare data from different anatomical locations to each other, then I would appreciate a sentence on their thoughts about how valid these comparisons are.

As explained in the initially submitted Material and Methods, we chose 50% of bone length as the standard location for diaphyseal parameter acquisition. However, two fossils did not include this region. So, in order to include them, we acquired anew the data for all specimens at the preserved regions as close as possible to 50% (which were in one case at 72% of bone length, and in the other at 35%). Only sections corresponding to strictly the same diaphyseal level were compared to each other. To make it clearer, we added a reminder to the revised Results. As stated in the figure legend (of the initial manuscript), Figure 3 does show mid-diaphyseal data. Please see also above our response to comment of Reviewer 1 on location of diaphyseal cross-sections.

18. Lines 342, 348, 354: I believe this is a typo: this should be "could be included", not "could have been included"?

Corrected in the RM.

Discussion 19. Page 16, lines 380-382: Has this been shown previously? If so, then requires a reference.

As explained in the Introduction, the present analysis is, to our knowledge, the first that combines in a single test both diaphyseal and trabecular parameters. The sentence was clarified in the RM.

20. Page 19: The most dramatic factors affecting bone structure during the life-course are likely age, gender and health status. This needs to be discussed.

This was added to the RM.

21. Lines 389-397: I don't understand the rationale behind proposing a size-corrected DA – it is a dimensionless parameter. Also I find the statement that DA scaled negatively in Doube et al, 2011 misleading: since (as the authors state) the p-value was >0.05 , surely one would conclude that DA does not scale?

It is true that DA is dimensionless. But that does not mean that it cannot be affected by scaling (in such a case the scaling coefficient denoting isometry would be 0; see for instance Ryan TM, Shaw CN. 2013 Trabecular bone microstructure scales allometrically in the primate humerus and femur. Proc R Soc B 280: 20130172.). Regarding Doube et al (2011)'s results about DA scaling, we agree that our initial sentence was misleading (which can be explained by the fact that Doube et al's results, as we understood them, are themselves ambiguous). We revised this sentence to replace this ambiguous reference.

Concluding remarks 22. I find the paper an interesting follow-up on the authors' group previous work on the xenarthran forelimb microarchitecture. The authors present some interesting thoughts on reconstructing anatomical function from as much information as is available in a clade that is not typically studied. I feel strongly that the authors should explicitly make clear that although bone responds to load, the nature of this response is a lot less predictable than the authors make it out to be currently. I feel the paper would further benefit from several clarifications as well as a more clear overall storyline to guide the reader through the ideas.

We hope to have addressed these points with the revision of the manuscript as detailed above.

Additional comments from a PCI recommender:

Suggestion to use Folivora rather than Tardigrada

We are in favour of using Tardigrada (which relates to the subclade name Eutardigrada). But we added a mention to the equally valid names Folivora and Phyllophaga in the RM.

I. 111. Suggestion to cite Pujos et al. 2012 Figure 4 Pujos, F., Gaudin, T. J., De Luliis, G., & Cartelle, C. (2012). Recent advances on variability, morpho-functional adaptations, dental terminology, and evolution of sloths. Journal of Mammalian Evolution, 19(3), 159-169.

Added to the RM.

I. 236. Suggestion to compare with the two last molecular publications for the age of the split between Bradypus & Choloepus (Slater et al. 2016; Delsuc et al.

2018) Slater, G. J., Cui, P., Forasiepi, A. M., Lenz, D., Tsangaras, K., Voirin, B., ... & Greenwood, A. D. (2016). Evolutionary relationships among extinct and extant sloths: the evidence of mitogenomes and retroviruses. *Genome biology and evolution*, 8(3), 607-621. Delsuc, F., Philippe, H., Tsagkogeorga, G., Simion, P., Tilak, M. K., Turon, X., ... & Douzery, E. J. (2018). A phylogenomic framework and timescale for comparative studies of tunicates. *BMC biology*, 16(1), 39. -1.419. Considering the strong uncertainties relative to the position of *Hapalops* (notably linked to strong changes in sloth phylogeny with the new results coming from aDNA), you should say on world on this as well, to moderate your results based on phylogenetically informed analysis. If you change the position of *Hapalops* (for example closer to *Bradypus*), does it change the result?

Reference to Slater et al. (2016) and Delsuc et al. (2018) were added to the relevant section of the RM (we assumed that it was Delsuc et al. (2018: Proc B 285: 20180214) that was meant). We agree that our assumptions regarding the phylogenetic affinities of *Hapalops*, as well as that of the other sloths, are likely to be strongly modified with the development of ancient DNA. The last version of our discriminant analysis for *Hapalops* does not recover a significant phylogenetic signal, so the test actually works as a traditional one (disregarding the phylogeny). We therefore do not expect much change for this analysis by altering the timetree to move *Hapalops* as sister-group to *Bradypus*. But we added a sentence to the RM to convey that sloths systematics is uncertain at the moment.

I. 462. You cite a manuscript under review.

The paper is now in press, we revised the manuscript accordingly.

I. 489. It is not only argued that Xenarthra would be one of the major clades of placental mammals; it is one of the 4 early diverging clades of extant placentals; you might cite molecular analyses for that as well.

Both suggestions are included in the RM.