

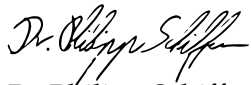
Cologne, September 6th 2023

Dear Dr Fernandez,

Many thanks for providing us with the second round of reviews by two experts in the field. We have gladly received their comments. Below we are replying to referee 1 in some detail and then provide a point-by-point reply to the minor comments the second referee, Dr Laumer, had. We hope with this second round of revision our manuscript will be fit for recommendation by you.

With kind regards,

on behalf of all authors,



Dr Philipp Schiffer

Ref 1:

However, my major concern is still that the authors present evidence of non-monophyly of Xenacoelomorpha as well as several of their results are perfectly in line with a non-monophyly of Xenacoelomorpha, while not providing direct evidence. All of these results are dismissed as artefacts without ever showing that they are artefacts. As mentioned in my previous review, the tip-to-root distances for Acoelomorpha (and especially Hofstenia) is not substantially longer than to other bilaterian species in the tree. Hence, there is no evidence for an increased rate of gene loss or gain in these lineages as the authors claim. Moreover, the authors provide no evidence or citation that the rate of evolution of gene gain or loss is correlated with the rate of evolution at the sequence level. The authors put this forward as an argument for dismissing all results concerning the non-monophyly of Xenacoelomorpha. The authors point out that increased substitution rates may result in problems that genes can get detected due to too much deviation from the sequences used as the queries (be it in a hmmer model or for blast searches). While in principle true, their results show no indication of this whatsoever as the acoelomorph species have similar or better scores than the xenoturbellid in their analyses. They can show such a reduction for orthonectids, which are also long-branched. Hence, this argument seems to be completely irrelevant for Acoelomorpha in this discussion and given the results presented. Finally, the authors argue that a series of phylogenomic studies have all supported the monophyly of Xenacoelomorpha. First, this support comes from a different source of data and hence the results of this study would be incongruent with the previous results. Hence, the results should be evenly discussed and considered and not just discarded as an artefact from the very beginning. Second, in a recent paper (Kapli, P., P. Natsidis, D. J. Leite, M. Fursman, N. Jeffrie, I. A. Rahman, H. Philippe, R. R. Copley and M. J. Telford (2021). "Lack of support for

Deuterostomia prompts reinterpretation of the first Bilateria." *Science Advances* 7(12): eabe2741) five of the authors of this paper (including the last author) advocated non-monophyly of Deuterostomia even though there is a plethora of studies supporting monophyly of Deuterostomia.

Hence, it appears inconsistent to advocate non-monophyly in this case, while to reject it outright here based on previous evidence. Taking this all together, the whole line of argument in this paper appears to dismiss all results, which are not in line with the cherished hypothesis (monophyletic Xenacoelomorpha as part of Deuterostomia) as artefacts, while all which are in line with it are taken at face value and put forward as strong support by extending the results from *Xenoturbella* to the whole of Xenacoelomorpha. This gives the impression of cherry-picking and safe-guarding the own hypotheses with ad hoc assumptions instead of given a fair and balanced view of the actual results presented.

In my previous review, this is what I suggested the author should provide. However, it is their paper and they will have to defend it in the future and it is not mine. If they want to have such a biased view published it is fine with me, but in my humble opinion it will weaken their whole line of argument. Why should others not claim that acoelomorphs show the actual situation of Xenacoelomorpha and that all results found in *Xenoturbella* can be dismissed as artefacts of some sort. This way of biased view can go both ways.

Reply to ref 1:

We thank the referee for their concern about the interpretation of our results. Clearly there are (at least) two possible trees relating *Xenoturbella* and the Acoelomorpha to each other and to other bilaterians. One is that they form a monophyletic group - and this is the solution that has been repeatedly supported by analyses of gene/protein sequences from multiple labs and using large data sets. This monophyletic relationship is also supported by signature peptides shared by *Xenoturbella* and the Acoelomorpha both in their caudal proteins and in one of their hox genes - these papers were cited in our revised manuscript. The alternative is the tree supported by our analysis of gene presence absence.

We feel that the weight of the evidence supports the monophyly of Xenacoelomorpha and we doubt that there will be many people familiar with the evidence and close to two decades of work on this topic who would disagree; although we agree that this is nevertheless formally possible. Our interpretation of the available evidence (and the fact that there is an alternative possibility which we clearly and openly discuss) is there for any reader to see and to disagree with if they wish. The reviewer makes a final point surrounding the question of deuterostome monophyly which we have a little difficulty following. The paper cited (*Science Advances* 7(12): eabe2741) is careful not to state that deuterostomes are non-monophyletic, rather this paper highlights the lack of strong support for monophyletic deuterostomes despite this clade being a feature of metazoan phylogenetic trees for more than a century. The support for monophyletic deuterostomes is seen in some but not all recent phylogenomic studies (compare this to the consistent support for

Xenacoelomorphs). The paper the referee cites shows that there is evidence that the support for monophyletic deuterostomes derives in part from LBA.

Ref 2, Chris Laumer, with point-by-point replies:

Having re-read this submission in full, I commend the authors on the considerable pains they have taken to improve the transparency and reproducibility of the manuscript (my major concern), and have also taken on board many of the small comments that I believe improve the interpretability of the figures. This is a compelling report and will go some considerable distance in informing the (sadly still ongoing) Nephrozoa vs Xenambulacraria debate - even if it is not decisive on the topology per se, I think the authors have used the genome and orthology analyses to compellingly argue against a “strong Nephrozoa”, and the detailed analyses of neuro-peptide complement, microRNAs, and gene content also should give pause to those who interpret Nephrozoa as a well-argued conclusion.

One thing I would like to see in the final version of this manuscript is a table displaying the accession numbers of all raw reads used in the study - those for assembly, scaffolding, and annotation. Naturally the final scaffolded assembly and annotation should also be made public, whether through NCBI/ENA (preferable) or Zenodo or some other mechanism, and accession numbers for these outputs should be cited in the paper. This will make it much easier for anyone else who wishes to re-assemble or re-use these data. Similarly, it would be nice to see public access to raw reads and assemblies made from other non-Xenoturbella specimens used in this paper (I am thinking of the Paratomella assembly used in the OrthoFinder analysis).

- We will make the accession numbers for individual datasets available in the final version of the manuscript as the reviewer suggests and do the same for the genome and annotation on ENA. At the time being these data can already be found under the X. bocki genome BioProject ID, which we state in the current version of the manuscript.
- We are now providing an excel file with the corresponding information for the species used in the OrthoFinder analysis. We point to this as Supplementary Information.
-

A few other small notes:

Typo on line 126 - presumably N50 is 8.5 kb, not 8,500 kb

- Fixed.

Discrepancy between text and table - is the final assembly span 111 Mb or 107.7 Mb?

- It is 111Mb accounting for unresolved bases/gaps (“N”s), please see the corresponding column in the table.

Please cite the SRA accession numbers of the RNA seq data you used as input to Braker.

- We have done so in the Methods section now. An accession for the RNA-Seq data was already available in the *Data availability* section since our first revision.

To clarify line 216: these are the orthologs present in the last common ancestor of all Metazoa? “all orthologs present in any bilaterian and any non-bilaterian” is a bit of a complicated way to say that.

- We are sorry for lack of clarity. The aim of this test was to know what genes must have been present in Urbilateria: from this we are able to count how many of these were missing in any bilaterian lineage. These urbilaterian genes were also, as the referee points out also present in Ureumetazoa although this is not of interest to us in this instance.
- We have reworded this as follows: “Using our phylogenomic matrix of gene presence/absence (see above) we identified all orthologs that could be detected both in Bilateria (in any bilaterian lineage) and in any non-bilaterian; ignoring horizontal gene transfer and other rare events, these genes must have existed in Urbilateria (and, of less interest to us, in Uremetazoa). The absence of any of these bilaterian genes in any lineage of Bilateria must therefore be explained by loss of the gene.

I thought I’d check this claim: “The model nematode *Caenorhabditis elegans* is ~81% complete for the same set.” and actually I got 78.5% for the *C.elg* N2 protein models downloaded of Ensembl Metazoa, with BUSCO5 and Metazoa odb10.

- We corrected our text to say “~79%”.

Access to the Zenodo data is embargoed for me so I cannot inspect it, but I do wonder, if it would be appropriate to also upload the draft *Paratomella rubra* genome & the raw reads used to derive it, to a public archive, since this is currently non-public data which was used to derive the OrthoFinder results cited in the text, if I understand right.

- We have uploaded the *P. rubra* raw reads (SAMN37209364), but think the genome is too fragmented to place it on any of the genome databases.
- We include it into the Zenodo container and hope to be able to sequence to better quality in the near future.

Line 471 - if acoels and Xenoturbella have similar levels of gene loss, then the long branches seen for acoels but not Xenoturbella in your gene presence/absence trees must be due to something other than gene loss. Furthermore, if this is just presence/absence data, it implies nothing about the rate of sequence substitution can be influencing the branch lengths. Surely therefore acoels have long branches in these trees only because they have gained many lineage-specific genes, no?

- We have included a note on this in the Discussion section now: “likely due to faster evolution, *gain of lineages specific genes*, and some degree of”

Line 513: I have to agree with the other reviewer - while you may or may not be correct in interpreting the acoel position in these analyses as an artifact, strictly, because of the non-monophyly of Xenacoelomorpha in this analysis, your analyses of gene content do support both Nephrozoa and Xenambulacraria, contradicting the text.

- We have remarked on this in the reply to the first reviewer.

Line 650: you say that you used the gVolante server for BUSCO5 assessment with the eukaryote USCO reference - contradicting the text earlier when you say you used a Metazoa database. Which is it?

- It was Metazoa. We now include percentages for both sets, Metazoa and Eukaryota

Supplementary figure 1 - Please describe how the blobplot was colored - was this diamond hits against uniprot, or blast against nr?

- We have done so. The former is the case.

Supplementary figure 2 - Which k-mer was chosen, what reads were used as inputs, and what software was used to perform k-mer counting? Please also show the full genome-scope plot, including details of the inferred genome size and fit of the model. From the glimpse I see here the WGS reads used for the raw assembly seem to be fairly low coverage, and it would help for a more clear-eyed interpretation of the primary assembly to understand any differences in the k-mer inferred genome size and the assembled genome-size.

- We have included the requested information, by adding the linear plot as second panel and stating the GenomeScope estimates and measurements.