


The logo for Peer Community In Genomics features a stylized network of blue and black dots and lines, with a circular pattern of black and white segments on the left side.

Peer Community In Genomics

Genomic idiosyncrasies of *Xenoturbella bocki*: morphologically simple yet genetically complex

Rosa Fernandez  based on peer reviews by **Christopher Laumer** and 1 anonymous reviewer

Philipp H. Schiffer, Paschalis Natsidis, Daniel J. Leite, Helen Robertson, François Lapraz, Ferdinand Marlétaz, Bastian Fromm, Liam Baudry, Fraser Simpson, Eirik Høyve, Anne-C. Zakrzewski, Paschalia Kapli, Katharina J. Hoff, Steven Mueller, Martial Marbouty, Heather Marlow, Richard R. Copley, Romain Koszul, Peter Sarkies, Maximilian J. Telford (2023) The slow evolving genome of the xenacoelomorph worm *Xenoturbella bocki*. bioRxiv, ver. 4, peer-reviewed and recommended by Peer Community in Genomics.

<https://doi.org/10.1101/2022.06.24.497508>

Submitted: 04 November 2022, Recommended: 22 November 2023

Cite this recommendation as:

Fernandez, R. (2023) Genomic idiosyncrasies of *Xenoturbella bocki*: morphologically simple yet genetically complex. *Peer Community in Genomics*, 100235. [10.24072/pci.genomics.100235](https://doi.org/10.24072/pci.genomics.100235)

Published: 22 November 2023

Copyright: This work is licensed under the Creative Commons Attribution 4.0 International License. To view a copy of this license, visit <https://creativecommons.org/licenses/by/4.0/>

Xenoturbella is a genus of morphologically simple bilaterians inhabiting benthic environments. Until very recently, only one species was known from the genus, *Xenoturbella bocki* Westblad 1949 [1]. Less than a decade ago, five more species were discovered (*X. churro*, *X. monstrosa*, *X. profunda*, *X. hollandorum* [2] and *X. japonica* [3]). These enigmatic animals lack an anus, a coelom, reproductive organs, nephrocytes and a centralized nervous system [1]. The systematic classification of the genus has substantially changed in the last decades, with first being considered as its own phylum (Xenoturbellida) and then being clustered together with acoels and nemertodermatids into the phylum Xenacoelomorpha [4,5]. The phylogenetic position of the xenacoelomorphs has been recalcitrant to resolution, with its position ranging from being the sister group to Nephrozoa (ie, protostomes and deuterostomes [6]) to the sister group to Ambulacraria (ie, Hemichordata and Echinodermata) in a clade called Xenambulacraria [4]. Recent studies based on expanded datasets and more refined analyses support either topology [7,8]. Either way, it is clear that additional studies on *Xenoturbella* could provide important insights into the origins of bilaterian traits such as the anus, the nephrons and the evolution of a centralized nervous system.

Small but mighty genome - In this work [9], the authors present the chromosome-level genome of *X. bocki*

- the first one for xenoturbellids - and explore their genomic idiosyncrasies in the context of other animal phyla. The first thing they discuss is the complexity of the genome, with *X. bocki* having a similar number of genes to other bilaterians (despite its small size of 111Mb), retained ancestral metazoan synteny, conserved clusters of Hox genes, largely complete signaling pathways and most bilaterian miRNAs present. This is not a surprise, though, as we know that the relationship between genomic and morphological complexity is far from straightforward - for instance, protist lineages closely related to animals share many gene families with us [10], and it is not the presence or absence of these gene families but their evolutionary dynamics what defines complexity in each animal phyla (eg [11]). However, the relationship between both is far from well-understood, and having a high-quality genome is the first crucial step towards a holistic understanding of genome evolution, allowing us to ask questions about how and when genes are regulated, how they interact in 3D space, or how their epigenetic landscape is shaped, for instance.

Xenacoelomorphs: deuterostomes or not? - The authors also discuss the phylogenetic position of xenacoelomorphs (including the newly generated high-quality genome of *X. bocki*) based on a gene presence/absence matrix. Although there is much more to be done to robustly assess the phylogenetic position of the phylum, these analyses represent a first attempt to investigate what the phylogeny looks like after the addition of the new high-quality data. The new analyses reflected once more the previously recovered phylogenies mentioned above, but this time with a twist: *X. bocki* was recovered as the sister group to echinoderms, yet acoels appeared as sister to all deuterostomes, hence not recovering Xenacoelomorpha as monophyletic. Thus, it is clear that much remains to be explored to disentangle the phylogenetic position of these mysterious lineages, where more sophisticated methodologies such as synteny-based orthology inference or models of evolution accounting for heterotachy probably have an important role to play.

In any case, we are approaching a qualitative jump in how we understand phylogenomics thanks to efforts derived from the availability of chromosome-level genome assemblies for a growing number of species. Exciting times are ahead for us, evolutionary biologists, to explore what high-quality genomes - in combination with multiomics datasets - will reveal about animal evolution. I am personally really looking forward to it.

References:

1. Westblad E. (1949). *Xenoturbella bocki* n.g., n.sp., a peculiar, primitive Turbellarian type. Arkiv för Zoologi 1, 3-29 (1949).
2. Rouse, G. W., Wilson, N. G., Carvajal, J. I. & Vrijenhoek, R. C. New deep-sea species of *Xenoturbella* and the position of Xenacoelomorpha. Nature 530, 94–97 (2016). <https://doi.org/10.1038/nature16545>
3. Nakano, H. et al. Correction to: A new species of *Xenoturbella* from the western Pacific Ocean and the evolution of *Xenoturbella*. BMC Evol. Biol. 18, 1–2 (2018). <https://doi.org/10.1186/s12862-018-1190-5>
4. Philippe, H. et al. Acoelomorph flatworms are deuterostomes related to *Xenoturbella*. Nature 470, 255–258 (2011). <https://doi.org/10.1038/nature09676>
5. Hejnol, A. et al. Assessing the root of bilaterian animals with scalable phylogenomic methods. Proc. Biol. Sci. 276, 4261–4270 (2009). <https://doi.org/10.1098/rspb.2009.0896>
6. Cannon, J. T. et al. Xenacoelomorpha is the sister group to Nephrozoa. Nature 530, 89–93 (2016). <https://doi.org/10.1038/nature16520>
7. Laumer, C. E. et al. Revisiting metazoan phylogeny with genomic sampling of all phyla. Proc. Biol. Sci. 286, 20190831 (2019). <https://doi.org/10.1098/rspb.2019.0831>

8. Philippe, H. et al. Mitigating anticipated effects of systematic errors supports sister-group relationship between Xenacoelomorpha and Ambulacraria. *Curr. Biol.* 29, 1818–1826.e6 (2019).

<https://doi.org/10.1016/j.cub.2019.04.009>

9. Schiffer, P. H., Natsidis, P., Leite D. J., Robertson, H., Lapraz, F., Marlétaz, F., Fromm, B., Baudry, L., Simpson, F., Høye, E., Zakrzewski, A-C., Kapli, P., Hoff, K. J., Mueller, S., Marbouty, M., Marlow, H., Copley, R. R., Koszul, R., Sarkies, P. & Telford, M. J. The slow evolving genome of the xenacoelomorph worm *Xenoturbella bocki*. *bioRxiv* (2023), ver. 4 peer-reviewed and recommended by Peer Community in Genomics. <https://doi.org/10.1101/2022.06.24.497508>

10. Suga, H. et al. The *Capsaspora* genome reveals a complex unicellular prehistory of animals. *Nat. Commun.* 4, 2325 (2013). <https://doi.org/10.1038/ncomms3325>

11. Fernández, R. & Gabaldón, T. Gene gain and loss across the metazoan tree of life. *Nat Ecol Evol* 4, 524–533 (2020). <https://doi.org/10.1038/s41559-019-1069-x>

Reviews

Evaluation round #2

DOI or URL of the preprint: <https://doi.org/10.1101/2022.06.24.497508>

Version of the preprint: 3

Authors' reply, 07 September 2023

[Download author's reply](#)

Decision by [Rosa Fernandez](#) , posted 21 July 2023, validated 21 July 2023

Revision needed

Dear Dr. Schiffer,

I would like to congratulate the authors for their efforts reviewing the manuscript. In light of the comments of both reviewers, I would like to invite you to revise your manuscript to continue addressing their comments and concerns. In particular, please take into consideration reviewer 1's comment on the lack of strong evidence supporting Xenacoelomorpha, together with making sure that accession numbers for all data generated or used in your analyses is properly acknowledge (after reviewer 2's comments).

I'm looking forward to reading a revised version of this piece of work.

Yours faithfully,

Rosa Fernández

Reviewed by anonymous reviewer 1, 10 July 2023

Schiffer et al addressed my comments and suggestions at least partially. Concerning Figure 4B, I would add in the legend what X, N and A means. Additionally, there are several typographic, grammatical and formatting errors in the text. At present, it appears like a test finished in a rush and the authors should carefully read it once more.

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 for 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.

Reviewed by [Christopher Laumer](#), 05 July 2023

[Download the review](#)

Evaluation round #1

DOI or URL of the preprint: <https://doi.org/10.1101/2022.06.24.497508>

Version of the preprint: 2

Authors' reply, 02 June 2023

[Download author's reply](#)

Decision by **Rosa Fernandez** , posted 10 January 2023, validated 10 January 2023

Invitation to revise your manuscript

Dear Dr. Schiffer,

Two reviewers have now assessed your manuscript. While they are both quite enthusiastic about the findings of this piece of work, they both raised major concerns that should be addressed before recommendation. In particular, they are concerned about the robustness of your results supporting the monophyly of Xenacoelomorpha, and the lack of sufficient detail ensuring transparency and reproducibility, among other minor issues. I encourage you to carefully address these concerns in a revised version of your manuscript.

Yours sincerely,

Rosa Fernández

Reviewed by anonymous reviewer 1, 15 December 2022

The paper «The slow evolving genome of the xenacoelomorph worm *Xenoturbella bocki*” by Schiffer et al. presents the first high-quality genome for a species of the genus *Xenoturbella*, which is the same as *Xenoturbellida*. The paper is very well written and relative easy to follow. The analytical steps are sound and state-of-the-art. This genome is large step forward towards our understanding of genome evolution in animals and will be a very useful tool for different kinds of research in the future. Most conclusions are well funded by the results, but I disagree with one conclusion the authors draw.

None of the results the authors present support the monophyly of Xenacoelomorpha (Acoelomorpha and *Xenoturbella*) and the phylogenetic analyses actually supports the non-monophyly. While the authors follow their phylogenetic results in concluding that *Xenoturbella* is placed within Deuterostomia, they disregard the non-monophyly of Xenacoelomorpha stating that Acoelomorpha are so fast evolving that this would mislead the phylogenetic reconstruction they conducted. However, their own results showed that the loss of genes in *Xenoturbella* and Acoelomorpha is not that different between the two groups. Moreover, there is no strong support yet that evolutionary rate at the sequence level is correlated with the rate of evolution at the synteny level. Finally, the tip-to-root distance for *Hofstenia* is not that different from the one of *Xenoturbella*. Hence, discarding the finding of non-monophyly of Xenacoelomorpha as an artefact appears like an ad hoc assumption in its strict meaning. As a consequence of this decision, the authors often treat findings related to the genome of *Xenoturbella* as relevant for the whole of Xenacoelomorpha, while the situation found in the genomes of Acoelomorpha are regarded as derived or data-insufficient given that they are of poorer quality.

My suggestion would be that the authors more clearly state the decision of favoring monophyly of Xenacoelomorpha is a subjective decision by the authors in contrast to their own results. Moreover, anytime they represent the findings for *Xenoturbella* as representative of Xenacoelomorpha, it should be indicated each time that this is done given a supposed monophyly of Xenacoelomorpha. In this way, it is might clear that this conclusion is not necessarily in agreement with their own phylogenetic results and might apply in the end only to *Xenoturbella* and not Xenacoelomorpha. In this way, the opinion of the authors is still present in the paper, but the presentation is a little bit more balanced.

In Figure 1A, please explain what the abbreviations mean.

The section at line 185, the first paragraph should not be part of this section as it deals with a different question. The paragraph deals with phylogenetic position of Xenacoelomorpha and does not contribute anything to the molecular toolkit. The remainder of the section is on the toolkit and that it is not different from other bilaterians. However, this does not add anything to the phylogenetic position. Hence, this should be

separated to avoid any implicit conclusions that one has to add something to the topic of the other, which they can't.

On lines 208-210, the authors state "Using our phylogenomic matrix of gene presence/absence (see above) we identified all orthologs present in any bilaterian and any non-bilaterian; these must have existed in the bilaterian ancestor." The conclusion is not a given. There are other options such hybridization, introgression, horizontal gene transfer and convergent evolution under similar selection pressures that could also result in this. Finally, wrong gene family detection due to poorly supported gene trees can be another reason. This should be mentioned.

On lines 255-258, the authors conclude that "Xenoturbella is, however, not significantly less complete when compared to other bilaterians considered to have low morphological complexity and which have been shown to have reduced gene content, such as *C. elegans*, the annelid parasite *Intoshia linei*, or the acoel *Hofstenia miamia*." The taxon sampling is very limited for this conclusion and many more complete bilaterian genomes are available. These should be mentioned here. Moreover, it should be explained how the author measure morphological complexity and what characters are the basis for this conclusion.

At line 266, the order of Hox genes in the brackets should reflect the order on the genome. This would highlight that typical colinearity of Hox genes seems not to be given.

In Figure 4b, it is not self-explanatory what the inset in bottom right corner explains? Please provide this explanation in the legend.

At line 356, it should be mentioned here that the ALG R is also present in the sea scallop.

In Figure 5b, it is not clear, which scaffold is the aberrant scaffold. I assume that it is c1896 as it is highlighted in red, but it should be explicitly stated to avoid any confusion.

Another small point is that the authors write concerning the phylogenetic reconstruction: "We calculated phylogenetic trees on these matrices using RevBayes (...), as described in ref 74,". This should be described in detail here as 74 runs two models, a reversible one and a Dollo one. It is not clear here, which one was used. This is especially relevant, if the Dollo analysis was applied. The Dollo criterion is a very strong assumption and hence it should be clearly stated if it had been used. If Dollo had been applied, I would advise to also conduct an analysis with the reversible model of 74 to test how the reconstruction performs without applying the very strong Dollo assumption.

Reviewed by [Christopher Laumer](#), 21 December 2022

Synopsis

Schiffer et al present a genome assembly, annotation, and comparative analysis of a representative of Xenoturbellida, perhaps the most evolutionarily interesting (and controversy-hounded) lineage of Bilateria, owing to its relatively simple gross morphology and uncertain phylogenetic position. They demonstrate a reasonably gene-space complete primary assembly and annotation of this small genome, and using HiC libraries scaffold roughly 3/5ths of the assembly into 18 linkage groups showing high levels of macrosynteny conservation with non-bilaterians and a representative deuterostome. A comprehensive orthology analysis shows, perhaps surprisingly regardless of the phylogenetic position of this lineage, a relatively Bilaterian-like gene content from the perspective of orthologue occupancy and signaling/transcription factor machinery, albeit showing a slightly higher than average loss of ancestral bilaterian orthologs (en par, I was surprised to see, with *Hofstenia miamia*, a representative of the acoelomorph sister lineage). A gene presence/absence phylogeny made with a different orthology declaration method and reduced taxon set shows strong support for a Xenambulacraria topology, even while splitting Xenacoelomorpha. Such an analysis is only possible with the whole-genome data presented here, and represents a refreshingly different and still somewhat novel approach to tackle this difficult phylogenetic problem to the familiar sequence alignment based inference methods, about which much has been published elsewhere. Numerous other "small" analyses (which I'm sure represent months of work in many cases), e.g. of miRNA content, neuropeptide complement, homeobox gene organization, phylostratigraphy, and symbiont genomics are presented, which shed light on many aspects

of xenoturbellidan biology - doubtless this manuscript will help solidify our understanding of this enigmatic lineage, and stimulate deeper study in some unexpected areas. The phylostratigraphically anomalous & sparsely methylated chromosome is particularly interesting.

There are a few apparent weaknesses of the manuscript. It's evident that these data were generated some time ago, and that the technologies used to generate the primary assembly are now basically obsolete - I'm sure that 1/3 of a Sequel II flow cell or a single MinION flow cell could generate a much more contiguous (and probably somewhat more complete) assembly of this genome with much less bioinformatic acrobatics these days. This said, I think the authors demonstrate convincingly that for the specific analyses shown in this paper, focusing on coding gene content and a birds' eye view of macrosynteny conservation, this assembly is adequate to the task at hand, and a reviewer shouldn't ask for more than that. This said, I would not present this as a "high quality genome" by today's standards - it is fundamentally a highly scaffolded Illumina genome which was just about contiguous enough to further scaffold to a pseudochromosomal stage with HiC data.

Obviously, one of the major uses of the genome will be in providing new evidence for the phylogenetic position of this lineage. I think the strength of the gene presence/absence phylogeny based on whole genomes assigned to OMA orthogroups speaks for itself, and I have no particular qualm with the authors' methods or interpretation of these results. However I did find it strange that this mode of phylogeny-building was not explored for the taxonomically much larger orthogroup assignment done using Orthofinder. True there can be failures to detect true gene presence in transcriptomes, and the acoel transcriptomes that exist vary quite a bit in quality, but the cynic in me did wonder whether such analyses were conducted and not presented because it yielded results incompatible with the authors' previous body of work on this phylogenetic problem. Some further justification of this decision, in any case, seems appropriate.

By far the most glaring problems with the manuscript are in its method section and overall transparency/reproducibility. Almost all of the primary data used to generate these results was not made available during review so that even basic sanity checks e.g. through a k-mer analysis of genome size & heterozygosity were not possible. Numerous basic reports e.g. on library quality and assembly statistics in various stages of the assembly pipeline were not presented. Important analyses are alluded to but not shown (e.g. blobplots, de novo transcriptome assembly statistics/completeness). Several clear factual errors are apparent (e.g. in the instrument used to generate the core assembly), and where both lab and bioinformatic protocols are remarked on, they are often presented with such a low level of detail as to forbid reproducibility. Indeed, many data types which were used for various small analyses (e.g. bisulfite sequencing, ONT sequencing) are not mentioned at all in the methods or supplement. I've given a fairly detailed account of where I see the absences in the notes below. For the most part, I have confidence in the quality of the datasets used to underpin this work, which was doubtless a lot of labor over many years, done the lab of a well established research leader in this field. I also do realize that these are *lots* of different experiments, and some of the data types are now no longer even on the market (e.g. TSLR). However, all published scientific literature should hold itself to a basic standard of transparency and reproducibility, which I would say this manuscript in its current form does not meet.

Signed,
Christopher Laumer

Detailed notes on introduction:

It seems the authors have made a choice not to cite any of the early molecular work plagued by contamination with molluscan gut contents. Follow up note: have the authors themselves done any screening for molluscan DNA?

From "line 70" - the authors refer to "a majority of studies" but cite only one (Cannon et al) - perhaps other citations are needed here?

From line 87 "The loss of...": A bit of a strange review - I'm not sure a barnacle or urochordate or neodermatan morphologist would characterize their study systems as morphologically simple. And is neoteny really a "new mode of living"? - I think of it as a hypothetical model of evolutionary change. I would have less issue with a

statement to the effect that major ecological transitions are often accompanied by major morphological shifts, including loss of “bodyplan” level features and organ systems.

Line 103 “The only Xenacoelomorpha genomes available...”: this is now out-of-date, with the preprint on the *Symsagittifera roscoffensis* genome, which is albeit very closely related to *Praesagittifera*. <https://www.biorxiv.org/content/10.1101/2022.08.27.505549v1.full>

Detailed notes on Results:

The difference in size between the primary assembly (121M) versus the final assembly (117M) suggests that very little sequence was removed by redundans as haplotypic duplication - is this correct? Was the genome relatively homozygous, e.g. as judged by kmer content?

I find the interpretation of false-negative orthology detection due to fast rates of sequence substitution leading to a splitting of Xenacoelomorpha in the p/a phylogeny quite credible, actually. There was an interesting paper recently published that looked at rates of false negative orthology detection and showed this to be a pervasive problem in taxonomic lineages that are poorly sampled and/or fast-evolving: <https://journals.plos.org/plosbiology/article?id=10.1371/journal.pbio.3000862>

I would be interested to know how many lineage-specific gene births and losses are recovered for *Xenoturbella*+*Ambulacraria* in this presence-absence analysis. Does the taxon-restricted gene set have any particular characteristics - e.g. average alignment length, compositional bias, distribution throughout the genome? If they are “perfectly normal” genes it strengthens the argument that this relationship is unlikely to be an artifact. I would be particularly interested in knowing if any of the “Xenambulacrarian” genes are particularly enriched on c1896, which was a very striking outcome of your analysis.

Line 207: is *Xenoturbella* slow-evolving or fast-evolving? The title says one thing, here another. Perhaps a little more clarity on what precisely is meant by this is worthwhile.

A comment on BUSCO completeness. One of the authors (PHS) was kind enough to share the genome annotation and assembly with me for use in a classroom module, before I was asked to review this manuscript. By my own hand, the BUSCO completeness of the genome annotation against *Metazoa odb10* in BUSCO5 (run in peptide mode) was:

C:82.5%[S:77.8%,D:4.7%],F:7.0%,M:10.5%,n:954

Which is substantially lower than the 90% quoted in the manuscript - is this down to a difference in database versions or has something else happened?

Also, in comparison to the figures from the genome annotation (whether 82.5% or 90%), I have separately done some re-analyses of publicly available RNA-seq data from *X. bocki* generated by other groups. My trinity assemblies give figures like:

SRR2681987 C:93.8%[S:22.6%,D:71.2%],F:2.2%,M:4.0%,n:954

SRR5760181 C:94.8%[S:27.5%,D:67.3%],F:1.7%,M:3.5%,n:954

This to my mind does indicate there’s some 40-50 metazoan BUSCOs present in adult RNA-seq data which are not represented in the genome annotation, potentially indicating some, although not a large level of, incompleteness of the genome assembly.

Line 211 “derived bilaterians” - In my view & in the views of many other systematists, no living species is more or less derived than any other - individual characters can be derived, but not whole organisms. I have a similar problem with reference to “early branching lineages” made elsewhere in the manuscript - from the perspective of cnidarians, for instance, mammals are a representative of an early-branching lineage. I am sure that the authors won’t disagree with these fundamental points, but the language that we sometimes use to describe these phenomena I think does bias our thinking towards a ladder-like view of evolution, and I would prefer to see these organismic comparisons made with less hypothesis-laden descriptors.

Very interesting that the bilaterian orthogroups have a similar occupancy for both *Xenoturbella* and the supposedly much faster evolving *Hofstenia*!

I am missing in Figure 3a data from the acoelomorphs incorporated into this analysis. Obviously you cannot show all 155 species in an easily readable way, but as a comparative point, it does seem essential to compare

the *Xenoturbella* species available to the acoelomorphs, particularly if you are trying to argue that this genome is especially faster or slower evolving than those from representatives its sister lineage.

For figures 3b/3c, are the trees at the bottom of the heatmaps dendrograms on the heatmap data, or are these schematic cladograms, and if so, from where do they originate?

I shall mostly refrain from commenting on the neuropeptide section as this is outside my expertise. I was curious, however - you see the K/R-RFP-K/R motif, which is reasonably compelling if anecdotal as a molecular synapomorphy, in *Xenoturbella* and in the ambulacrarians, but not in the nemertodermatid transcriptomes - is it also present in the reasonably complete transcriptome from *X. profunda*? Or any of the acoels?

On ancestral linkage groups. Indeed the conservation of macrosynteny visible with other metazoans is an impressive feature of this assembly, and a compelling demonstration of the success of this paper. I think it would help improve the readability of this manuscript if you could e.g. put a bold-line box around some of the clade-specific linkage groups you discuss in the oxford plots. Also, I don't understand the argumentation around "prebilaterian" linkage groups. Neither the Nephrozoa hypothesis nor the Xenambulacraria hypothesis posits that Xenacoelomorpha are non-bilaterians - why would the absence of a eumetazoan plesiomorphy say anything decisive about either hypothesis?

On the anomalous small chromosome: super interesting result, and indeed perhaps the start of some really interesting *Xenoturbella*-specific biology. I wonder if this will be seen to occur in any acoelomorph genomes going forward. I would be very keen to see if gene tree topologies/delta-likelihoods of orthologs occurring on this chromosome are any different on average to those occurring elsewhere in the genome - for instance indicating a potential horizontal origin (albeit you don't see a large signature of this in the global analysis of HGT). One other thing: I couldn't really find anything in the figures corresponding to the synteny with the *E. muelleri* scaffolds described in the text - could you make this clearer? Indeed I don't see the c1896 labelled on the dot plot with *E. muelleri*.

Detailed notes on Discussion:

The discussion of "intermediate" genomic traits such as miRNA counts and linkage group organization feels a bit phenetic to me. Surely all of the traits mentioned could be analyzed cladistically, in a search for synapomorphies. Simply intermediate numbers of various character states shouldn't be compelling on their own.

Is the relatively canonical gene content of *Xenoturbella* informative either way on the Xenambulacraria vs Nephrozoa debate? I'm not sure I agree that it is. I think for instance of the *Dimorphilus* genome which was recently published, showing many features we would expect a typical annelid genome to have, despite the highly reduced body size and morphological simplification of this lineage. This to me shows a decoupling between a birds' eye view of genome biology and morphology. So indeed, while the relatively bilaterian-gene rich genome of *Xenoturbella* is consistent with the Xenambulacraria hypothesis, it's not *inconsistent* with the Nephrozoa hypothesis either. I do like your phrasing of the "strong" Nephrozoa hypothesis not being supported - this does imply, however, that a "weak" Nephrozoa hypothesis is possible (presumably meaning a Nephrozoa true tree topology but little obvious genomic "pre-bilaterian-ness" as one might naively think if one interprets xenacoelomorph morphology as primitively simple and gene content as predictive of morphology). Indeed, you say as much in the final paragraph of the discussion - I think this measured reading is appropriate and commendable.

Sentence beginning line 461: I had thought that the long branch lengths seen for acoels in your presence/absence trees would indicate a high rate of acoelomorph-specific gene births, rather than a high rate of loss? Is it possible to disentangle these?

Another possibility about the anomalous chromosome: could this be a germline restricted chromosome? We do expect that these should have younger genes on average, and these are also usually small. Does it show a different level of average coverage in the raw reads to other chromosomes?

Detailed notes on Methods:

Which phenol-chloroform protocols and Qiagen kits were used for DNA extraction? In the results it's asserted

that HMW gDNA was extracted - how was this ascertained/QC'd?

It's concerning to me that the authors state a 2x250 bp read format was used on the HiSeq 4000, as this platform does not offer that read length. Perhaps it was a HiSeq 2500 2x250 rapid run?

It would be good to see one of the blobplots referred to, to convince the reader that this really is an uncontaminated genome assembly, despite the efforts to starve the specimens before extraction. This is a part of the tree of life for which few close references exist and it can be tricky sometimes to judge the source of contaminants from blobplots on such species - perhaps better to show rather than tell.

I am missing here some basic statistics (perhaps best shown in a table) on these assemblies during various points in the process, e.g. right after SPAdes, after redundans, after BADGER. The authors cite a scaffold N50 of 60 kb before HiC scaffolding, but what is the contig N50 before any scaffolding?

Indeed, the authors refer to mate-pair libraries but do not give any details on the protocols used to generate these datasets, the size of the mate pairs, QC statistics...

Redundans should be cited.

I was unable to find a link to the raw reads (except for two HiC datasets) or assemblies used in this paper. I was hoping to do some basic analyses, e.g. kmer spectra, just to cross-check for instance that the assembled genome size matches the kmer estimated size, and to determine what proportion of kmers in the reads were not represented in the assembly. Without such data (which could have been uploaded to SRA and embargoed for public release) it's difficult to fully review this manuscript. I will note that the genomescope analyses I made of the two HiC datasets were somewhat concerning, with no visible spectra outside an error distribution - perhaps these are low-diversity libraries, or highly contaminated libraries?

Some concerns about the HiC protocol and data presented here. Fixing a whole animal vs fixing cryohomogenized tissue is likely to lead to poor results from autolysis as the fixative penetrates large volumes of tissue. There's no indication that the DpnII enzyme was heat-killed before proceeding to fill-in. Numerous volumes used (for formaldehyde and SDS, for example), and enzyme details (which "ligase"? what manufacturer?) are missing. It's not clear what protocol was used to prepare the extracted 3C DNA into an Illumina library, or how the biotin selection was performed. And most concerning of all, I can't really see any QC data on this library - at very least, the authors should be showing the pair-length distribution and the contact heatmaps which have become standard in the field, so that readers can judge how strong the evidence for the chromosome scaffolding is.

I think, as instaGRAAL is a published method, it's not necessary to explain its algorithm in detail here - just the parameters that were used to run it.

The protocols used for RNA extraction and cDNA library preparation should be specified in enough detail for another lab to reproduce this work. It would also be good to see some rough statistics on the Trinity assembly, so readers can judge its completeness and contiguity. Again, I could not find any RNA-seq reads used in this study uploaded to the SRA.

The authors refer to additional single-cell transcriptome data - if these were used in the annotation of this genome, surely the experiments used to derive these should also be described in the methods section and deposited into public databases?

Question: during setup for orthology inference, for the species for which RNA-seq data only were used as input, were *only* those genes with positive hits against UniProt/Pfam retained in the protein prediction, or was this simply used to improve the sensitivity of the predictions? I am wondering if this pipeline might exclude novel taxon-specific orthologs with no sequence similarity to existing databases.

I don't see any problem per se in the way that the gene presence/absence phylogenies were generated, but I am curious why the OMA algorithm, and apparently a separate species set, was employed in this while the Orthofinder analysis should also in principle be well-suited to this kind of analysis. Do the results differ with a larger taxon sample?

In the homeobox section, ONT reads are mentioned for the first time. There's no information given in the manuscript about the volume and quality of these data, and how they were generated. I also find it strange

that these were used only in the context of homeodomain-containing contig analysis - why not also incorporate them into the primary SPAdes assembly?

Line 751 "We extracted a highly contiguous..." - how was this extraction performed bioinformatically?

Line 752 - As I understand it, LINKS is a scaffolder, not a polisher.

BUSCO is mentioned - including re-analyses of public data such as the Hofstenia genome - but the parameters/database versions used to run this software seem not to have been reported.

Similarly: there are some results on methylation reported in the supplement, but no mention is made of how these results were obtained - was this bisulfite sequencing? If so, how were these libraries generated and these analyses performed?