

We thank the recommender and the reviewers for his/her comments on the manuscript, to which we provide the following answers.

Round #1

by Wirulda Pootakbam, 2020-08-09 08:47

Manuscript: <https://www.biorxiv.org/content/10.1101/2020.05.25.110734v1>

Please revise the manuscript

This study provides an important resource to the field. The authors did a thorough study on the vertebrate microbiota; however, there are issues raised by reviewers that need to be addressed before I can recommend this preprint. I would like to ask the authors to carefully go through each reviewer's comments and try to address them.

Reviews

Reviewed by anonymous reviewer, 2020-07-15 09:54

Scalvenzi and colleagues provide a comprehensive analysis of the *Xenopus*' microbiota along its distinct developmental stages. Important questions such as the variation of the bacterial community composition along these stages, along the gut, but also the impact of diet on *Xenopus*' microbiota, and the source of transmission of the community were investigated. An interesting parallel is made between mammals and *Xenopus* microbiota, with some major lineages and functions being conserved. I believe that this thorough piece of work could prove important for the communities working on vertebrate's microbiota, on *Xenopus* evo-devo, or host-microbiota associations, and establish a reference for future investigations.

The article reads generally well, and conclusions drawn seem well-supported by the data.

Just a small point on my reviewer's experience: I have to admit that it was quite hard to find figures (some were duplicated in the main text PDF file – provoking some confusion at 1st sight), sup figures, and sup tables. Figures and tables are not all numbered in the combined PDF files, and there seems to be some issues with sup figures numbering (see below). Putting the legend next to each figure - and a title by each table would have been a great help to ease up the reading. Please also note for next submission that all sup files could have been submitted independently to bioRxiv.

Answer : We apologize for these confusions and we prepared the revision according to your suggestions, with legends next to each figure and table.

Here follow specific points of discussion:

- 1) I was a bit surprised that archaea were not mentioned at any point of the study. They are increasingly recognized as being part of mammals microbiota (humans, apes, cows...)—yet in relatively little abundance (see for instance Koskinen 2017, mBio). As a parallel is drawn between mammal/vertebrate microbiota and that of *Xenopus*, I was wondering whether they had popped out at some point in the metagenomic/ metatranscriptomic analyses. Unfortunately, it has been shown that universal primers often used for 16S amplification miss a huge portion of archaeal diversity and that others should be used (see Raymann 2017, mSphere), but metagenomic/metatranscriptomic study could have revealed them.

Answer : We thank you to raise this point. We looked for Archaea but did not find compelling evidence for their presence in metagenomic or metatranscriptomic reads. We used kraken2/braken and kaiju and found low levels of reads assigned to Archaea but these turned out as false positives upon a closer inspection using megablast on SILVA or the NCBI non-redundant nucleotide databases. We discuss this limitation in the manuscript and put it in the agenda as an avenue for future research.

2) lines 121-124: “we analyzed the distribution of bacterial populations by relative size (forward scatter) and nucleic acid fluorescence, we found that the samples from young tadpoles differed markedly from those of mature tadpoles in their cytometric profiles (Figure 1A).” I’m not much familiar with flow cytometry techniques... but could these results translate into distributions by cell size on e.g. barplots? To me, this seems easier to compare and grasp differences between distributions than by visualizing the type of graph on Fig 1A.

Answer : Thank you for this remark. The bivariate dot plot, also called a scattergram, shown in Figure 1 is used widely for cytometric results and this is why we used it ¹. The rationale for using this visualization is that it translates the simultaneous measurement of DNA quantity (measured by propidium iodide fluorescence) and relative cell size (forward scattering). We reanalyzed the cytometric data and modified Figure 1 to make it more easily understandable. In addition, we provide a simplified representation to help the interpretation of such results in supplementary Figure 1. We also clarified the text in the corresponding results section.

I don’t find evident the differences in communities in terms of cell size between the earlier tadpole stage and the others, except that the number of cells is much lower for the former. Could for instance “population A” of the earlier stage correspond to “population C” at the NF56 stage?

Answer : Thank you for these comments. As written before, we have modified Figure 1 to make it more easily understandable and point clearly our interpretations. We also provide a simplified representation to help the interpretation of these results in supplementary Figure 1. We agree that certain bacterial populations correspond from one stage to another, like 50-3/56-4/60-3 or Juv-4/Adu-3. It is less obvious for the other populations.

Are the differences between distributions significant? What statistical test was performed? A related question: maybe I missed those, but on which grounds were defined the “clusters”/ “bacterial populations” based of flow cytometry results presented in Fig 1A? I don’t see any details provided either in Methods or in Sup Methods.

¹ Corinne Ferrier Maurice et Peter James Turnbaugh, « Quantifying and Identifying the Active and Damaged Subsets of Indigenous Microbial Communities », *Methods in Enzymology* 531 (2013): 91-107, <https://doi.org/10.1016/B978-0-12-407863-5.00005-8>.

Answer : We used only descriptive statistics on the distribution of these cell populations, and provide a graphical representation in supplementary figure 1. We performed a statistical analysis only on the number of bacteria quantified at each life stage as described in the manuscript and documented on the R scripts (Kruskal-Wallis test, ANOVA and pairwise t-test). We identified the bacterial populations by visual inspection based on the normalized scatter plots shown (200000 events).

4) Fig. 1A: what are the cells at around 10^2 forward scattered, and 10^1 PI? Do they correspond to dead cells? Otherwise, couldn't they form another population than the "A" one?

Answer : Thank you for these questions. These were not considered to be cells, but cellular debris and noise¹.

5) Line 122: please clarify for non FACS specialists, that you are talking about cell sizes.

Answer : Thank you for this comment. We modified the text to make it clearer and use the term relative cell size.

6) I think the results on Fig S3C could call for a small comment on inter-individual variations at the same developmental stage, given than samples are then pooled for analyses as shown on e.g. Fig. 2A.

Answer : Thank you for your remark. We modified the text to highlight the inter-individual variations: "In most cases, inter-individual variations were coherent with the global changes observed with the previous or the next developmental stage, and this is why we pooled our observations to analyse them at the level of gross developmental stages."

7) If I may, I have some suggestions for Figure 2BCD: maybe could different colours distinguish between developmental stages? Also, for homogeneity-sake, could the same terms always designate the same developmental stages? "Premet" instead of "tadpole" for instance, on all panels? Maybe the legend could display the developmental stages' in the "right order" to help non-specialists follow?

Answer : Thank you for spotting these issues. We corrected them. We swapped Fig_S3 and Figure_2 according to a similar remark made by the other reviewer.

8) Line 206 - ... : it is not alpha diversity that is displayed on Fig 2B, but a measure of phylogenetic diversity (Faith' PD index). It could be appropriate to precise which metrics exactly were used when mentioning the results, and upon 1st appearance, slip a few words about the exact meaning of each metrics.

Answer : We thank you for having spotted this mistake here. We modified this sentence and introduced the metrics used.

9) Line 211: which Sup Table? Aren't they all numbered? Line 238, 255, 263,...: same comment

Answer : We checked and corrected the references to supplementary material provided as additional files.

10) Please explain what are exactly the metrics used for "community structure" and "community membership" on Fig. 3? Why are the distance' types used different between Fig 2C and Fig 3B?

Answer : We apologize for this lack of clarity due to a mistake in Figure 2 legend. In Figure 2, we made an ordination using unweighted unifracs distances to analyse community membership (Fig2C), and using PhilR metric for community composition (Fig2D). Unfortunately, we made a mistake in the legend and referred the right panel of Fig2C as community structure instead of community composition. In Figure 3, we made an ordination using unweighted unifracs distances to analyse community membership, and using weighted unifracs distance for community structure as indicated in the figure legend.

We used different names to reflect the use of different metrics between Fig2C (community composition, PhilR metric) and Fig 3B (community structure, weighted unifracs). We explored a variety of metrics to evaluate their impact on the ordination results (Jaccard, unifracs, Bray-Curtis, weighted unifracs, philR) as illustrated in https://npollet.github.io/metatetard/xpdev_analyse_ordination_phyloseq.html. We showed the ordination results that provided the best explanation of the data in Figure 2C and Figure 3B (ie explaining more variance), and we acknowledge that this may lead to confusion. We modified Figure 3 and used PhilR metric for community composition to simplify the reading.

11) Line 256: “Feces and skin microbiomes were clearly separated from the other samples, but not between them.” Maybe would be worth mentioning that these are also the ones that vary the most among a same organ?

Answer : We agree and rephrased this sentence accordingly :

Feces and skin microbiomes were characterized by their largest variability and were well separated along the x-axis in the ordination analysis.

12) Line 329, 331, 412: “SES MPD”, “SES MNTD”, “RPK”, please explain acronym upon 1st appearance.

Answer : We apologize for this jargon and corrected this lack of acronym explanation.

13) Fig. S8B: What is exactly the metrics used? what is measured? “w”.

Answer : We apologize for not mentioning in Fig. S8B that the text above the box-plot graph is the result of a Wilcoxon signed-rank test and that W is the value computed by this test. We have modified the legend accordingly.

14) How much gDNA and RNA material was used for prepping the metagenomic/metatranscriptomic sequencing libraries?

Answer : We provided 4 µg of gDNA and 11 µg of total RNA to our service provider, the BGI. We do not know the exact amount used for the library construction, but the amount of DNA and RNA was not limiting and not in the lower range of nucleic acid amounts for sequencing.

15) Programs used to assemble and map the reads of the metagenomic analyses could be briefly mentioned along Results.

Answer : Thank you for this remark. We modified the main text to mention the softwares used for the assembly (ray meta, megahit, metaspades, idba_ud) and for the functional annotations (prokka, minpath, ipath).

Line 554: the general method to assess the source of microbial communities is only mentioned in Discussion – this could have been introduced earlier.

Answer : We modified the result section to introduce the method used to evaluate the source of microbial communities :

In a general manner, methods used could be briefly listed along main text so that the reader understands where the results come from.

Answer : Thank you for this comment. We modified the text to briefly outline the methods used to derive the results.

16) Please check sup figures numbering: Figure S9 and S10 rather S7 and S8. Please also check that all sup figures are cited in main text – I am not sure this was the case.

Answer : We apologize for these numbering errors. We checked that all additional files are cited (and this was the case).

17) The assembly statistics are not given in a very extensive manner, but given the median size of assembled contigs (~800bp), it seems quite sub-optimal. I was wondering what was the proportion of assembled contigs predicted as being part of a CDS? This could give an idea of the proportion of assembled reads that has been analysed.

Answer : Thank you for your remark. We improved our assembly by running different assemblers (ray, megahit, metaSPADES and idba_ud) using different parameters to finally select an assembly with an N50 of 6100 bp. The ray assemblies consistently provided assemblies with lower N50 but with other merits. We provide detailed metrics of our assembly in the text, in the extended materials and methods and in supplementary table 3-additional file 13).

18) Line 405: it is unclear to me what scaffolds were analysed: all assembled ones? or the 25 longest only? Please clarify (results for 10 are shown on fig S9-S10).

Answer : Thank you for your remark. We clarified our results on the assembly, and to answer directly your question we analysed all scaffolds.

19) It is said line 403-404 that each of the 10 (or 25? unclear, see above) longest scaffolds “seems to be derived from a different genome”: what exactly goes in this direction? I don’t think this can be assessed from results shown on Fig S9 and S10 (GC% and reads coverage along scaffolds). Could it be possible to assign taxonomy to the longest scaffolds? It is said on line 681 that the BGI provided taxonomic affiliation for metagenomic data, is that correct?

Answer : As mentioned in our answer to your question 17, we updated this section and performed new assemblies. We performed taxonomic affiliation of metagenomic and metatranscriptomic reads using MATAM² and phyloFlash³.

20) Fig. S10: reads mapping is quite uneven for the scaffold at the bottom, with a shift towards lower coverages toward the right of the contig, making this scaffold look suspicious. Could the authors comment on that?

Answer : As mentioned in our answer to your question 17, we updated this section and performed new assemblies.

21) Line 672: “July 6th to 9th” from which year?

Answer : We apologize for this reference mistake, this has been corrected.

22) Line 676: the name is “Silva”.

Answer : This has been corrected.

23) Line 691: please check again sup fig numbers.

Answer : We triple checked references to all additional files.

Reviewed by Vanessa Marcelino, 2020-07-04 00:49

The manuscript from Scalvenzi and colleagues reports a comprehensive study on the microbiome of *Xenopus*, an important model organism. The authors report significant changes in microbiome composition during developmental stages. The study also addressed the impacts of diet and parental transmission of microbes. The majority of the study was based on amplicon sequencing, with metagenomics and metatranscriptomics data from a subset of samples constituting a nice addition to the study. Overall, this thorough study provides an important resource to the field. There are however several issues that need to be addressed before I can recommend this paper for acceptance.

The potential effect of confounders and sample size is unclear There are multiple co-variates and it is important to be upfront about those. For example: the authors sequenced 16S amplicons from both RNA and DNA, why?

Answer : Thank you for raising this general point. We agree that the study of microbiomes is still replete of confounding aspects, as pointed out in several review papers (Pollock et al., 2018; Bharti and Grimm, 2019; McLaren et al., 2019). We strived to do our best using the existing methodologies, and it is clear that more studies will be needed to refine our observations, reproduce

² Pierre Pericard et al., « MATAM: Reconstruction of Phylogenetic Marker Genes from Short Sequencing Reads in Metagenomes », *Bioinformatics* 34, n° 4 (15 février 2018): 585-91, <https://doi.org/10.1093/bioinformatics/btx644>.

³ Harald R. Gruber-Vodicka, Brandon K. B. Seah, et Elmar Pruesse, « PhyloFlash: Rapid Small-Subunit rRNA Profiling and Targeted Assembly from Metagenomes », *MSystems* 5, n° 5 (27 octobre 2020), <https://doi.org/10.1128/mSystems.00920-20>.

them using alternative technologies and more powerful tools. We added a paragraph in the discussion section to discuss limitations of our study.

We used 16S rRNA gene sequencing from both DNA and RNA to infer bacterial viability⁴. Since tadpoles are microfeeders, we reasoned that working with DNA extracted from tadpole's guts could be under the main influence of bacteria from the « environment » and used as food. At the same time, we know that most gut microbiome data are issued from DNA studies, and we wanted to be able to compare our finding with that of other studies. For a proper inference of the physiological roles played by bacteria, we reasoned that it would be more relevant to work with RNA⁴. Finally, we decided to have some data points using both RNA and DNA, so that we could evaluate our results and those from the literature in a broader context.

Are the statistical results still significant when this is taken into account?

Answer : Statistically speaking, data have been processed independently. A fair statistical analysis of the difference between DNA and RNA as starting material would require an extraction of both DNA and RNA from the very same samples, and this was not our case. Thus, we refrained from such an analysis.

The genetic material was preserved in ethanol for some samples, while fresh DNA/RNA was used for others. I would like to see evidence that this differential treatment does not affect the conclusions of the paper.

Answer : We used fresh material for all RNA extraction because it is one of the simplest and best method for obtaining pure and non-degraded RNA. We used ethanol preserved tissues for all DNA extractions. We used this methodology because we processed more samples for DNA extraction. Since it has been shown that ethanol preservation is adequate for microbiome studies⁵, we decided to preserve tissues in at least ten volumes of absolute ethanol. In addition, working on such ethanol-preserved tissue will give the opportunity to other laboratories to use the same strategy, and ultimately to perform comparisons with tissues sampled from wild populations.

We do not mix the analysis of results obtained using DNA or RNA to draw the main conclusions of our paper. The results on « *Xenopus* gut microbial diversity during development and metamorphosis », « *Xenopus* microbiota across several gut compartments » and « *Xenopus* microbiome transmission » were obtained from extracted DNA. The results on « Activity of the *Xenopus* tadpole gut microbiome at the onset of feeding » were, as its name entails, obtained from extracted RNA. « Diet shapes *Xenopus* tadpole's gut active communities » also (we added « active » to underscore the difference). Finally, DNA and RNA for shotgun metagenome and Metatranscriptome were from freshly sampled tissues. Throughout the manuscript, we used a clear

⁴ Steven J. Blazewicz et al., « Evaluating rRNA as an Indicator of Microbial Activity in Environmental Communities: Limitations and Uses », *The ISME Journal* 7, n° 11 (novembre 2013): 2061-68, <https://doi.org/10.1038/ismej.2013.102>.

⁵ Emily Vogtmann et al., « Comparison of Collection Methods for Fecal Samples in Microbiome Studies », *American Journal of Epidemiology* 185, n° 2 (15 janvier 2017): 115-23, <https://doi.org/10.1093/aje/kww177>; Doris Vandeputte et al., « Practical Considerations for Large-Scale Gut Microbiome Studies », *FEMS Microbiology Reviews* 41, n° Supplement_1 (1 août 2017): S154-67, <https://doi.org/10.1093/femsre/fux027>; Se Jin Song et al., « Preservation Methods Differ in Fecal Microbiome Stability, Affecting Suitability for Field Studies », *mSystems* 1, n° 3 (28 juin 2016), <https://doi.org/10.1128/mSystems.00021-16>; Eric A. Franzosa et al., « Relating the Metatranscriptome and Metagenome of the Human Gut », *Proceedings of the National Academy of Sciences of the United States of America* 111, n° 22 (3 juin 2014): E2329-2338, <https://doi.org/10.1073/pnas.1319284111>.

vocabulary to distinguish these two approaches: « 16S rRNA gene profiling” versus “16S rRNA sequencing on whole RNA extracts”.

Given that the Methods section is at the end of the manuscript, it is important to mention number of replicates for each experiment while describing the results.

Answer: Thank you for this remark. We modified the text to clearly state the number of samples for each experiment.

The metagenomics and metatranscriptomics aspects were performed with only 5 samples, all from the same developmental stage, and therefore it does not contribute to “analyse the succession of microbial communities and their activities across different body habitats” as the abstract (and other parts of the manuscript) suggest. This small sample size is only mentioned in the methods section, at the end of the manuscript. The manuscript would benefit from being upfront about this limitation.

Answer : Thank you for this remark. We apologize for this confusion. We did not claim the use of a single approach to analyse the succession of microbial communities, and wrote initially in the background section of the abstract « We used different approaches including flow cytometry, 16s rDNA gene metabarcoding, metagenomic and metatranscriptomic to analyze the succession of microbial communities and their activities across different body habitats of *Xenopus tropicalis*”.

We clarified in the abstract what methodology was used and for what purpose

Gender bias: Line 63 reads “... these results were challenged by Eugène Wollmann and his wife, who finally observed...” Referring to a female scientist as just someone’s wife is sexist. Cite both first names or none. Also – I think the reference is Wollman and Wollman, 1915 (rather than Wollman 1913, but please check). Likewise, in line 503, the word “man” can be replaced by the more neutral word “humans”.

Answer : We corrected this. The references are correct, the 1913 paper deals with the experiment on gnotobiotic tadpoles, and the 1915 paper entitled « microbes in tadpole food » (‘Les microbes dans l’alimentation des têtards’) deals with the nutrition.

The manuscript needs **English proofreading.**

Answer : The manuscript was proof-read by two different English-proficient colleagues.

I also have questions regarding the interpretation of some of the results:

X. tropicalis microbiome transmission: The authors conclude that the skin of the parents is one of the main drivers of microbial transmission, but they did not sample an important source of microbes – the surrounding water. Considering that the adult frog skin microbiome is likely to be influenced by the surrounding water, it is possible that their conclusion about the parental transmission is circumstantial, and that the surrounding environment is the single most important contributor.

Answer : Thank you for raising this point. We agree that the environment, e.g. mainly water, is the most important contributor. We apologize that we did not make it more clear, but actually this is what we wrote :

“In five out of the eight crosses, more than 70% of the bacteria found in the eggs were not from feces or the skin of the parents.” And in the last sentence of our result section on the transmission : « In conclusion, we found that the main drivers of egg bacterial communities were the environment and the skin of the parents, with the feces being only minor contributors.”

We modified this last sentence to clarify this point and avoid any misunderstanding, and we added details of our methodology.

To avoid being subject to an important contribution from free-living bacterias from the aquarium water we took some precautions. Upon sampling the skin, animals were abundantly rinsed with sterile amphibian PBS and swabbed for several minutes using a cotton swab.

Xenopus gut microbial diversity during development and metamorphosis: L.148 – 149:

How were these OTUs defined (which % similarity)?

We defined OTUs using the FROGS pipeline ⁶. Quoting this paper : « The advantages of FROGS are that it relies on Swarm ⁷ and its adaptive sequence agglomeration rather than on a global similarity threshold, combined with a rigorous chimera removal step and the explicit consideration of conflicting affiliations. ». Specifically, clustering is first done with the aggregation parameter $d = 1$ and a second pass is performed on the seeds of previous clusters with $d = 3$. According to the authors, this equates to <5% divergence at the genus level when using the 16S V3V4 region.

L 149- 150: The rarefaction curve is far from reaching a plateau. In fact, the overall sequencing depth is low. The authors should be upfront about this limitation.

Answer : Thank you for this remark. We agree and are aware that our dataset has limitations, mostly due to a restricted capacity to budget further sequencing. We preferred having more biological replicates, even if we have also been limited here. We used different strategies, DNA and RNA-based metabarcoding, because we are convinced that it provides a better view of the microbiome ⁸. Limitations have been nicely evidenced by the studies of the Microbiome Quality Control Project Consortium ⁹. As a direct consequence of your remark, we re-estimated the species coverage in our samples and we modified our analysis by using breakaway, a recent statistical approaches developed to account for measurement errors in diversity analysis ¹⁰.

Technically speaking our graphical representations of the rarefaction curves were not clear enough, and indeed a plateau was reached for most samples (see modified Figure S2, all rarefaction plots are at the same scale). In addition, we show species accumulation curves that argue for the

⁶ Frédéric Escudie et al., « FROGS: Find, Rapidly, OTUs with Galaxy Solution », *Bioinformatics* 34, n° 8 (15 avril 2018): 1287-94, <https://doi.org/10.1093/bioinformatics/btx791>.

⁷ Frédéric Mahé et al., « Swarm v2: highly-scalable and high-resolution amplicon clustering », *PeerJ* 3 (10 décembre 2015), <https://doi.org/10.7717/peerj.1420>.

⁸ Franzosa et al., « Relating the Metatranscriptome and Metagenome of the Human Gut ».

⁹ Rashmi Sinha et al., « Assessment of Variation in Microbial Community Amplicon Sequencing by the Microbiome Quality Control (MBQC) Project Consortium », *Nature Biotechnology* 35, n° 11 (novembre 2017): 1077-86, <https://doi.org/10.1038/nbt.3981>; Michael R. McLaren, Amy D. Willis, et Benjamin J. Callahan, « Consistent and

Correctable Bias in Metagenomic Sequencing Experiments », *ELife* 8 (10 2019), <https://doi.org/10.7554/eLife.46923>.

¹⁰ Amy D. Willis, « Rarefaction, Alpha Diversity, and Statistics », *Frontiers in Microbiology* 10 (2019),

<https://doi.org/10.3389/fmicb.2019.02407>; Amy Willis et John Bunge, « Estimating Diversity via Frequency Ratios », *Biometrics* 71, n° 4 (2015): 1042-49, <https://doi.org/10.1111/biom.12332>.

identification of the major components of the bacterial microbiome since they reach a plateau (Figure S2).

The confounding effects due to an heterogeneity of sampling depth are the most worrisome. A restricted sampling is likely to lead to an erroneous and inexact species richness estimation. This mostly comes from rare species, i.e singletons distribution across samples. We explored this point and did not found clear limitations (shown in https://npollet.github.io/metatetard/xpall_rarefaction_phyloseq.html and https://npollet.github.io/metatetard/xpdev_breakaway.html).

In most gut microbiome studies, or microbial-rich environments, most singletons are dropped out early on during the bioinformatic steps of read clustering and chimera removal, distorting species richness estimates ¹¹. In fact, there is a correlation between the level of diversity in a dataset and the precision and exactitude of richness estimation (discussed in ¹². Overall in our analysis we focused on the most abundant and the most prevalent OTUs, and this was our major route to limit this confounding factor. In our project we filtered out clusters with abundances below 0.005% abundance (all samples combined) as recommended by previous studies. In our revised manuscript, we estimated richness and other diversity metrics using statistical modelling approaches that provide error measurement. The take home message is that it did not change our conclusions.

Xenopus tropicalis gut microbiota genes catalog The conclusion that the *Xenopus* gut microbiome is similar to the human microbiome is most likely an artefact. The majority of known bacterial genes are from human-derived bacteria, therefore the results observed here are not surprising. I invite the authors to compare the available gut metagenomes of any other vertebrate (especially the ones raised in captivity). The conclusion that the *Xenopus* gut microbiome is similar to the human gut microbiome is only relevant if the authors can show that this is not the case for other vertebrates.

Answer : A comparative analysis of vertebrate microbiomes is beyond the scope of our paper. We acknowledge your point that the human microbial gene catalog is the source of potential artifacts and we dropped these aspects that would require more extensive comparisons between various vertebrate species.

L 402-403: I don't see how you can make the conclusion that the scaffolds come from different genomes (you mean organisms?) based on coverage.

Answer : We made a completely new assembly and included a binning strategy to infer the genomic origin of scaffolds. We used the software DASTool that includes diverse strategies to group scaffolds that are likely derived from the same bacterial genome¹³.

¹¹ Willis et Bunge, « Estimating Diversity via Frequency Ratios ».

¹² Amy Willis, John Bunge, et Thea Whitman, « Improved Detection of Changes in Species Richness in High Diversity Microbial Communities », *Journal of the Royal Statistical Society: Series C (Applied Statistics)* 66, n° 5 (2017): 963-77, <https://doi.org/10.1111/rssc.12206>.

¹³ Christian M. K. Sieber et al., « Recovery of Genomes from Metagenomes via a Dereplication, Aggregation and Scoring Strategy », *Nature Microbiology* 3, n° 7 (juillet 2018): 836-43, <https://doi.org/10.1038/s41564-018-0171-1>.

Metabolic profile of the *Xenopus* gut microbiota: Were vertebrate and human-contaminant sequences removed from the analyses? This is essential to ensure that the observed metabolites were derived from the microbiome.

Answer : This is a good point, and yes we used BWA-MEM to map all raw reads to the nuclear and the mitochondrial *Xenopus tropicalis* genome, and filtered out all paired reads with at least one mapped read. We modified the materials and methods accordingly.

It is not clear to me how the metabolic map connecting the host and the microbiome pathways was performed, and the interpretation of these results is even less clear. The paper states that “Our data highlighted the important capacity of the microbiota to complement the metabolic pathways of the *Xenopus* genome” (L. 573 – 574). I think this an over statement. The data does not indicate if the metabolites produced by the microbiota can be used by the host (or vice versa). Supplementary figure S11 (the metabolic map) is really impressive, but not mentioned anywhere in the main text.

Answer : We modified the materials and methods to explain the methodology used to infer metabolic pathways. We describe in more details the analysis of the pathways provided by the microbiome regarding carbohydrate breakdown, short-chain fatty acids biosynthesis, nitrogen recycling, essential amino-acids synthesis and B-vitamins biosynthesis.

The discussion around bile acids is interesting, it seems highly speculative but the authors acknowledge that this is just one possible interpretation of the results.

Answer : We are happy to have conveyed our interest on this speculation.

Evolution This study does not provide insights in the ‘evolution of microbiomes’ as suggested by the title, and it also does not help to understand the “evolutionary associations between amphibians and their microbiota.” as stated in the conclusion. There is no evidence for an evolutionary association between host and microbes here. Perhaps succession is a better word.

Answer : We changed the title so that it reflects more accurately the content of our study.

Specific comments:

Ethics approval: Please mention the issuing institution of this ethics approval.

Answer : We modified the text to name the institution providing the ethics approval : Direction Départementale des Services Vétérinaires de l’Essonne.

The R scripts (L. 678) are a plus, really nicely done. Please add the link in the “availability of data and material”.

Answer : Thank you for this remark. We added the link in the section “availability of data and material”.

Line 681 - 682: “We used the taxonomic assignment results provided by the BGI for the metagenomic analysis.”. Please provide more details, this is not reproducible.

Answer : Thank you for pointing that out. We changed this section accordingly.

The bias in microbiome composition that is caused by keeping animals in captivity is only superficially addressed in the last paragraph of the discussion, without any qualitative or quantitative assessment. How does the microbiome composition observed here compares with the one observed in wild anurans? I suggest performing this comparison and adding this information in the results section.

Answer : Unfortunately, we do not have access to enough data obtained from *Xenopus tropicalis* tadpoles sampled in the wild. In principle, we could extract from public databases the gut microbiome data obtained from sampling wild-caught anurans and go on with whatever high-level analysis. Yet, we refrained performing such a metadata analysis that would be very biased by design : different methods to collect specimens, to produce metabarcoding data, different species, different samples (feces vs gut content). We added a sentence to raise this limitation of our study in the discussion.

Supplementary Materials:

Supplementary tables are hard to find and what I think is Sup table 1 seems to be incomplete (it does not show number of sequences obtained per sample as mentioned in L. 668).

Answer : Thank you for pointing that out. We included a header to the supplementary tables to find them more easily in the pdf file containing all supplementary materials. We corrected Supplementary table 1 so that it includes the number of paired reads per sample.

Did you perform rRNA depletion during library preparation?

Answer : No, we did not perform rRNA depletion during library preparation for metatranscriptomic sequencing. This was on purpose to obtain rRNA sequences without an amplification bias.

DNA extraction from animals fixed in ethanol - what was this used for? Metagenomics?

Answer : Thank you for this comment. We fixed tissues and whole tadpoles using ethanol to preserve them before DNA extraction. We have clarified this section that aimed to document the methods used for tissue sampling. We included a novel paragraph entitled « Tissue sampling » in the materials and methods of the manuscript where we explain the procedures for sampling gut tissues, skin, feces, eggs.

Metagenome assembly: the text states “We used the Genotoul hardware and software services (<http://bioinfo.genotoul.fr>) to perform the assembly and used other high memory usage software.” Please specify which other software were used. Sentences like this make the whole manuscript sound like a coarse draft.

Answer : Thank you for pointing that out. We changed this sentence to acknowledge the support of Genotoul services. It is always a pleasure to read an insightful comment such as “Sentences like this make the whole manuscript sound like a coarse draft.”

Figure S3 is also really good and make the whole study much clearer. I would try to reduce the text in the results section (which is rather long) and replace it with this figure.

Answer : Thank you for this remark and for this suggestion. We swapped Figure 2 and Figure S3 and modified Figure 2. We reduced the text accordingly.