December 30th 2019,

Dear PCI team and recommender,

We wish to submit the revised version of our manuscript now entitled 'A flexible pipeline combining clustering and correction tools for prokaryotic and eukaryotic metabarcoding' to PCI. We addressed all comments by reviewers, and clarified some points in view of the comments.

In particular, as highlighted by reviewer 2 and yourself, we have provided further clarification of methods (including more explanations regarding mock communities) as well as a better justification of the bioinformatic parameter choices, both in the methods section and in the result section by including several *d*-values used for clustering with swarm v2.

We detail here below the changes made, and hope this revised version will be considered as suitable, and are available for further clarification if needed.

Sincerely,

Sophie Arnaud-Haond, on behalf of co-authors

**Institut français de Recherche
pour l'Exploitation de la Mer**

Etablissement public à caractère
industriel et commercial

Station Ifremer de Sète
Boulevard Jean Monnet
CS 30 171
34200 Sète
France
+33 (0)4 99 57 32 00
Siège social
1625 route de Sainte-Anne
CS 10070
29280 Plouzané
France

http://www.ifremer.fr

Institut Français de Recherche pour l'Exploitation de la Mer (IFREMER)
MARBEC, Bd Jean Monnet, BP 171, 34200 Sète - France
Tel: +33(0)4 99 57 32 61 sarnaud@ifremer.fr

Rev#1 Tiago Pereira

R1.1: *The study by Brandt et al. (A flexible pipeline combining bioinformatic correction tools for prokaryotic and eukaryotic metabarcoding) brings new insights into data analysis of metabarcoding datasets covering both prokaryotes and eukaryotes as well as mitochondrial (e.g. COI) and nuclear genes (e.g. 18S and 16S), particularly with the inclusion and testing of new methods/bioinformatic tools (e.g. DADA2 and LULU). It is an interesting and well-written paper, likely to be very useful to many biologists/ecologists dealing with these types of datasets. The reviewer has done minor comments/changes in the pdf file (see attachment).*

Comment and actions 1.1: we included all minor changes in the revised file, except:

- We did not remove "tube" in tube cores (lines…), as it is important to differentiate between tube cores and blade cores that are distinct equipment allowing to sample different volumes of sediment.
- We defined BLAST in the Materials and Methods section rather than in the abstract.

Additionally, the authors should consider the following major points:

R1.2: *Expected relative abundance: multi copy nature of rRNA genes, PCR bias, etc., might confound our expectations. How close is it good enough?*

This was the aim of the mock communities, i.e. to approach as much as possible a correspondence one species-one OTU. We clarified this by changing the order of the results and discussion, to clarify this line of thought. We also recovered a higher abundance of reads for those species with higher relative DNA concentration in the mock communities; yet, the *in vitro* admixture of DNA does not mimic the diversity of species with distinct number of cells, repeats of ribosomal sequences or mitochondria… For insects (Pinol et al. 2019 Mol. Ecol.) or fishes (Pont et al. 2018 Scientific Reports, Toshiaki et al. 2017, Mol. Eco. Res.), correlations between biomass and read abundance could be established using very specific primer pairs. However, our global approach using "universal" primers to encompass the broadest possible range of metazoan diversity does not allow such approach. This is the reason why we prefer not to use the information about read numbers and to stick to presence/absence for metazoans, while for microbial diversity we adopted the classical approach present in the literature accounting for the number of reads in each ASV.

R1.3*: Intragenomic/intraspecific polymorphism: is this a real problem? Can we alleviate by using phylogenetic methods?*

Comment 1.3: The split of a given species sequence into several ASVs counts among the most important pitfalls of metabarcoding. As this "demultiplication" is not comparable across taxa (as seen in newly added Fig. S2), the taxonomic compositions of samples will not be reflecting species diversity but a combination of genetic diversity and divergence, which is not the aim of a biodiversity inventory. With the mock samples, we show that clustering ASVs into OTUs alleviates this problem for most metazoan species and LULU curation further improves accuracy,

2

yet we also see that the problem is not entirely solved: some species still produce more than one OTU. It has thus to be kept in mind that an exact 1 to 1 species/OTUs relationship can still not be expected with the tools available today. The help of phylogenetic methods is limited in Illumina Sequencing due to its restriction to short DNA fragments, but can certainly be valuable with the development of longer fragment sequencing using alternative sequencing technologies (PacBio, Nanopore) and DNA enrichment methods (such as capture by hybridization or long-range PCR).

Action 1.3: we clarified the problem generated by intraspecific polymorphism in the introduction (lines 81-107) and in the discussion (lines 501-519)

R1.4: General trend/patterns: although the different methods produced different results (e.g. alpha/beta diversity), how strongly did they impact the overall pattern?

Comment 1.4: We clarified in the manuscript that the impact is high (and significant) on alpha diversity, and thus on biodiversity inventories (depending on the pipeline, some OTUs are lost, others are split). For beta diversity, we show that the patterns of community structures are qualitatively consistent across pipelines, when d-values are low (Fig. 2). Although LULU curation leads to significantly different communities in 18S V1-V2, as revealed by the significant PERMANOVA results, this does not change the overall spatial patterns observed and may be explained by different abundances of spurious OTUs in some major phyla (Arthropods, Cnidarians, Nematodes…).

Action 1.4: We added a figure illustrating the changes in ASVs/OTUs detected per taxon in each pipeline in the Supplementary material (Fig. S2). We also added a figure (Fig. 2) illustrating beta-diversity patterns to clarify this point in the result section. We clarified the text in the discussion to explain the effects of pipeline parameters on alpha vs beta diversity, lines 572-579 and 588-596.

R1.5: In the pipeline, what seems to be the crucial step (e.g. clustering methods/thresholds or taxonomic assignment) in order to produce reliable/accurate findings with respect to biodiversity and ecological patterns?

Comment 1.5: We found that both steps are crucial as they both modify the results obtained, but in different ways. Clustering and LULU curation improve overall ecological patterns as they influence alpha and, to a lower extent, beta diversities (see comment above). Contrastingly, taxonomic assignments affect the qualitative value of biodiversity inventories, and the precision in the description of communities at a finer grain taxonomy level. The pipeline we present here aims at providing global inventories. As seen in Fig. 3, neither BLAST nor RDP provide satisfactory taxonomic assignments of deep-sea datasets using available public databases, at least not for COI. Carrying out a two-step assignment, first on a global database as done here, and then a second level assignment, phylum by phylum, with databases specific to the species and ecosystems investigated will be of value to improve biodiversity inventories, yet such databases still need to be built, in particular for the deep sea.

Action 1.5: We modified the structure of the manuscript to better separate and detail those two issues: we first address the improvement obtained for "numerical ecology" using mock and true samples. We then look at taxonomic assignment quality with

both BLAST and RDP for the three loci, and underline the need for a two-step assignment using more specific databases in the discussion section, which reads:

"*At present, if accurate taxonomic assignments are sought while using universal primers, we advocate assigning taxonomy in two steps: first, using BLAST and a large database including all phyla amplifiable by the primer set, extracting the clusters belonging to the groups of interest, then re-assigning taxonomy to these target taxa using RDP and a smaller, taxon-specific database.*"

R#2: Anonymous reviewer

R2.1: *My review of the language will therefore be very short and does not include typos (but there are some!). I would suggest, though, to re-structure the order of some paragraphs, which might improve the reading experience of the manuscript even further.*

Action 2.1: We went through the annotated pdf to incorporate annotations and corrections. We also restructured the manuscript for more clarity, in order to address the reviewer's comment to "*disentangle the results of the mock community approach from the results of the 'true' samples. One possibility is to restrict oneself first to the mock community results, because they allow for setting the further results in a context*"

The results were re-ordered as follows:
1. Alpha diversity in mock communities
2. Alpha diversity in natural samples
3. Beta diversity patterns in natural samples
4. Taxonomic assignment quality

Major concerns:

R2.2: *Stefanni et al., 2018; Multi-marker metabarcoding approach to study mesozooplankton at basin scale. Scientific Reports 8:12085). Stefanni et al., made some different choices regarding their bioinformatic pipeline, but their work and results should at least be discussed in the context of the current manuscript here.*

Comment and Action 2.2: We thank the reviewer for the mentioning of this reference. Similarly to our approach, Stefanni et al. combined COI and 18S (although they used 18S-V9) and showed a strong difference in alpha diversity levels obtained depending on clustering threshold, but also depending on taxonomic similarity threshold used to assign metazoan sequences (cut-off value of 80% for COI and 86% for 18S). We could not comparatively evaluate the effect of LULU-curation as we did not find a mention to the parameters chosen in the manuscript or associated sup. Mat. We thus now cite this work in the discussion as another example of study showing the improvements that can be gained through more conservative bioinformatic pipelines, yet also emphasizing the need for flexible assignment thresholds in less well-known communities such as deep-sea ones (lines 612-630).

Indeed, Stefanni et al.'s study focuses on shallow-water coastal environments and mesozooplankton assemblages (mostly macrofauna), which are much better

represented in public databases. Moreover, many phylogenies of invertebrates (including shallow water genera) show high levels of species divergence, even within genera. This is the case for example for Eunicidae (Zanol et al., 2010, and associated GenBank data), where most species show a minimum of nearly 20% (and up to more than 30%) divergence from any other species at COI. Similarly, phylogenies of deep-sea taxa such as barnacles (Herrera et al, 2015) or Alvinocarid shrimp (Shank et al., 1999) showed that some species had COI sequences diverging more than 20% from any other species present in molecular databases. Considering that Alvinocarid shrimp count among well-represented deep-sea taxa, this clearly predicts a large amount of low homologies in understudied ecosystems and taxa such as deep-sea ones (even more so when including smaller meiofauna). These data show the need for much more flexible thresholds for taxonomic assignment (acknowledging and accepting associated uncertainties) than the ones applied to well-known terrestrial or even coastal taxa. As we explain below (comment 2.6), when we tried using a 96% assignment threshold for COI, most of our OTUs were discarded. This is also highlighted in Fig. 3, where BLAST assignments for COI on a marine-only database rarely go beyond 90%.

The second major difference in Stefanni et al.'s approach is the merging of the COI and 18S data for both alpha and beta diversity assessment. This is very interesting, and something we actually planned to do in our ecological analysis of the deep-sea sediment samples. As COI and 18S are very complementary in terms of taxa targeted (highlighted in Fig. S2), subsampling each dataset for its "best targeted phyla" and subsequently combining both seems to be a very efficient way to produce comprehensive and non-redundant biodiversity inventories. We have incorporated this line of thought in the discussion (lines 600-615).

R2.3*: Using LULU in combination with DADA2 was originally tested by Frøslev et al., 2017 on plant data. I am not convinced that simply applying the same combination on metazoan, eukaryotic and prokaryotic data is enough for a study that proposes a 'flexible pipeline combining bioinformatic correction tools', because neither tool was developed by the authors, nor is said combination a novel idea of the authors. Maybe the authors refer to the combination of DADA2 and Swarm for being the proposed novel flexible pipeline. If that is what they are aiming at, they may want to consider putting the combination of DADA2 and swarm (and LULU) in the focus. Momentarily, it reads as the focus is on DADA2 and LULU.*

Comment 2.3: LULU was validated on plant ITS data, but invertebrates and metazoans are very different both in terms of markers and genetic divergence patterns: the authors of LULU recommend in their discussion that the LULU parameters will have to be evaluated for different taxonomic compartments, and ideally on mock communities. This is, among other things, what we aim to do in this work. Our goal, i.e. to have a flexible pipeline allowing more accurate inventories of eukaryotes and prokaryotes, is precisely to 1) evaluate the need for clustering in these taxonomic compartments, currently debated in the metabarcoding community (Callahan et al 2017), and 2) evaluate the effectiveness of LULU for metazoans. We thus do not propose the same pipeline for prokaryotes and eukaryotes

Actions 2.3: Following the reviewer's advice, we now insist on and clarify the advantages of swarm when dealing with metazoans. In particular, we modified the structure of the introduction to clarify the advantage of swarm v2. We made clearer that different pipeline parameters are advisable depending on the barcode-targeted taxonomic compartment at stake.

R2.4: Several parameters were chosen in the bioinformatic pipeline that are currently not justified in the text. The most prominent example is Swarm's d value, which is set to 4 for 18S data, 6 for COI data and 1 for 16S data (lines 261-262). I am aware of only few studies that do not use Swarm's default of d=1, most likely because the results become harder to interpret. Allowing a difference of one nucleotide between two sequences in one OTU can easily be justified by naturally occurring sequence variation or artificially introduced sequencing errors. Every value beyond d=1 is harder to justify and may be just as arbitrary as the clustering thresholds the authors try to avoid. In fact, I was surprised that the authors use the avoidance of arbitrary sequence similarity clustering thresholds as an argument for Swarm (lines 54-55 and 113-115), but then try to set d to a value that mimics a 1% sequence divergence threshold, which is just the invers of a 99% sequence similarity threshold (lines 349-351).

Comment 2.4: We understand that the selection of clustering thresholds is arbitrary and the investigator's choice based on pilot experiments. This is what we attempted doing here. Swarm with d=1 does not mean that all sequences within an OTU will all have 1 nucleotide divergence, as swarm makes networks of sequences at *d* (iterative linking) and then breaks weak links to produces its OTUs. Similarly, with *d=4-6*, the OTUs produced by swarm v2 will not only contain sequences of 1% divergence. In addition, several studies on metazoans have used *d* values of up to 13 (Macias-Hernandez et al 2018, Siegenthaler et al. 2019, Kemp et al. 2019). We thus in fact had tested several d values for swarm, but did not provide the detail of those tests in the previous manuscript.

Action 2.4: We have included in the work a selection of *d* values that cover the range used in the literature (depending on the marker, *d=1 to d=13*). We present the results of these different *d* values in the alpha diversity section of the MS, and present effects on beta-diversity on a selection of *d* values and their corresponding LULU-curated version.

R2.5: Swarm OTUs clustered with a different d value are pooled and analyzed in the same context. In my opinion, OTUs that are analyzed together should always be treated as similar as possible. I suppose the size of the 18S V1/V2 region is nearly as long as the 16S V4/V5 region; why were then so different thresholds chosen for the clustering of the respective OTUs? The authors need to justify these decisions and if they cannot come up with scientifically sound justifications, they should consider sticking to those values that are justifiable.

Comment 2.5: The rationale behind these different choices is as follows: the most employed biological species concepts do not apply to agamospecies (lacking sex) such as prokaryotes. Microbiologists thus usually favour avoiding arbitrary clustering and keep all genetic diversity as a possible *proxy* for divergence and/or ecological niche adaptation. However, clustering still makes sense for prokaryotes when aiming at studying associations with other taxa (such as symbiotic relationships between prokaryotic and eukaryotic OTUs), and above all for metazoans, as it avoids biasing biodiversity inventories by putting more weight on species with higher levels of intraspecific polymorphism. Finally, if OTUs are produced, clustering thresholds need to be adjusted to the marker gene chosen, i.e. its intrinsic mutation rates, and taxonomic resolution level. Consequently, COI data needs to be clustered at higher *d-values* as 18S, just because it displays higher polymorphism, and prokaryotic 16S data, when clustered, will be minimally so.

Action 2.5: We have included the test of a range of d-values for the three loci in order to improve comparability. We clarified the importance of adjusting the *d* values depending on marker gene and taxonomic compartment targeted in the discussion section (lines 541-570).

> *R2.6: Other more or less arbitrary values for which I found no explanation or justification were the maximum error rate for primer removal in CUTADAPT (lines 231-232), the truncation length, maximum expected error rates (line 243) as well as the minimum overlap for paired-end assembly (line 247) in DADA2, the very low identity (70%) cutoff for BLAST (line 254) and the minimum match values for LULU (line 280). All of these parameters have a severe effect on downstream data processing and ultimately on the results. Maybe the authors chose the values for a good reason or they followed default values from the literature. But without further explanations, the readers cannot understand their decisions and I would not recommend using a bioinformatic pipeline that does not inform about such important steps.*

Comment and Action 2.6: We have added more explanations of the different parameter choices in the methods section:

- For cutadapt, the sentence reads: "*Briefly, the script uses cutadapt v1.18 to detect and remove primers, while separating forward and reverse reads in each paired sequence file to produce two pairs of sequence files per sample named R1F/R2R and R2F/R1R. Cutadapt parameters (Table S3) were set to require an overlap over the full length of the primer (default: 3 nt), with 2-4 nt mismatches allowed for ribosomal loci, and 7 nt mismatches allowed for COI (default: 10%)*"
- For *filterAndTrim* in DADA2, the sentence reads: "*Reads were filtered and trimmed with the filterAndTrim function and all reads containing ambiguous bases removed. The parameters were set based on tutorial recommendations and trimming lengths were adjusted based on sequence quality profiles, so that Q-scores remained above 30 (truncLen at 220 for 18S and 16S, 200 for COI, maxEE at 2, truncQ at 11, maxN at 0).*"
- For the *learnErrors* function and the merging step in DADA2, we used the default values, and specified this in the text.
- For taxonomic assignment based on BLAST, we use a (relatively low) minimum hit identity of 70% for two reasons: first, we have tried implementing higher taxonomic thresholds, but this lead to the loss of most OTUs as databases are currently under-represented in deep-sea taxa. Second, we have observed inter-specific mitochondrial DNA divergence levels of up to 33% within a same deep-sea polychaete genus (*Eunice* sp). We have added a sentence to reflect these thoughts: "*A second taxonomic assignment method was optionally implemented in the pipeline, allowing assigning ASVs using BLAST+ (v2.6.0) based on minimum similarity and minimum coverage (-perc_identity 70 and –qcov_hsp 80). An initial test implementing Blastn+ to assign taxonomy only to the COI dataset using a 96% percent identity threshold ejected the majority of the OTUs, reducing the power the analyses. Given this result, as well as observed inter-specific mitochondrial DNA divergence levels of up to 33% within a same*

*deep-sea polychaete genus (Eunice sp), and our interest in the
identities of multiple, largely unknown taxa and poorly characterized
communities, more stringent BLAST thresholds were not implemented
at this stage."*

R2.7: *In abstract and introduction, the authors make a point about the importance of
multiple marker metabarcoding approaches. However, they conclude that DADA2 is
not fit for analyzing metabarcoding datasets of metazoan organisms (lines 504-507).
In contrast to this finding, there are at least two publications that analyzed metazoan
metabarcoding datasets with DADA2 and did not report the problems presented by
the authors here. One of the publications used the 18S V4 marker region and was
cited by the authors (Xiong & Zhan 2018), the other publication used the 18S V9
marker region and was not cited by the authors (Leff et al., 2018; Predicting the
structure of soil communities from plant community taxonomy, phylogeny, and traits.
ISME Journal 12:1794-1805). These studies show that i) the conclusions about
metazoan metabarcoding data drawn by the authors on base of the COI region
cannot be generalized to all gene regions and ii) the authors may have targeted a
less suited gene region for their approach. In any case the results of the current study
should be discussed in the context of these previous studies.*

Comment 2.7: We are sorry about this unclear statement as we by no means wanted
to state that COI is not fit for DADA2 (or the other way round), but rather that
pipeline parameters need to be carefully adjusted to the marker used, and its intrinsic
properties. COI, evolving much faster than ribosomal gene regions such as 18S,
displays much higher intraspecific variation and thus requires more stringent analysis
parameters.

Although presenting limitations (see Galtier et al. 2009), COI still remains the
standard barcode for metazoans as it offers the best species-level resolution (as seen
in the mock samples) and reference databases, including also advantages for curation
such as its protein coding nature (Andujar et al., 2018). Using only ribosomal
markers has been shown to seriously underestimate biodiversity (Tang et al. 2012),
thus combining mitochondrial and ribosomal DNA is the best approach to get the
most comprehensive biodiversity inventory for metazoans. Rather than disregarding
COI, we have thus tried to adapt the pipeline to the marker genes chosen.

Action 2.7: We now clarified in the manuscript that the impact of clustering or not is
significant on alpha diversity, and thus on biodiversity inventories (i.e. species
detections). However, for beta diversity, we show that the patterns of community
structures are qualitatively consistent across pipelines, when *d*-values are low (Fig.
2), which is in line with findings reported in *Xiong & Zhan 2018 and Leff et al.,
2018*. We have modified part 3.3 of the discussion to explain our results in light of
existing literature, and better highlighting the differences between loci.

R2.8: *Although I admit that it is a tedious topic, I was also surprised about the
author's choice of the 18S V1/V2 region instead of the more commonly used V4 or
V9 region. Can the authors please comment on why V1/V2 was chosen? Much more
reference data seems to be available for V4 and V9. Since correct taxonomic
assignments were an important topic in the current study, using a marker gene for
which more reference data is available would have been beneficial for the authors'
study design.*

Comment 2.8: Our primary goal was to evaluate pipeline effects on metazoan datasets; we chose the 18S V1-V2 region targeted by the Sinniger primers (Sinniger et al. 2016) as these primers are modified to preferentially target metazoans and are also well represented in the curated reference database Silva. Preliminary results on 18S V4 and V9 data that are also produced in our project show that these regions target many more protists over metazoans, and better results are obtained for metazoans using 18S-V1 and COI.

Minor comments:

R2.9: Two sentences I struggled the most with:
'*As metabarcoding with multiple markers, spanning several branches of the tree of life is becoming more accessible, bioinformatic pipelines need to accommodate both micro- and macro biologists.*' (lines 2-4).
'*The results also confirm an important variation in the amplification success across taxa (Bhadury et al., 2006; Carugati, Corinaldesi, Dell'Anno, & Danovaro, 2015), supporting the present approach combining nuclear and mitochondrial markers to achieve more comprehensive biodiversity inventories (Cowart et al., 2015; Drummond et al., 2015; Zhan, Bailey, Heath, & Macisaac, 2014).*' (lines 542-546).
Could you please rephrase to make it clearer to the reader what you want to express?

Action 2.9: We re-phrased these sentences, they now read:

"*With sequencing costs decreasing, multiple-marker metabarcoding, spanning several branches of the tree of life is becoming more accessible. However, bioinformatic approaches need to adjust to taxonomic compartment as well as marker gene specificities.*"

and

"*Our results also support approaches combining nuclear and mitochondrial markers to achieve more comprehensive biodiversity inventories (Cowart et al. 2015; Drummond et al. 2015; Zhan et al. 2014). Indeed strong differences exist in amplification success among taxa (Bhadury et al. 2006; Carugati et al. 2015), exemplified by nematodes, which are well detected with 18S but not with COI (Bucklin et al. 2011).*"

R2.10: The numbering of the manuscript sections is askew. Introduction should be '1', but Methods ended up being '1' and so on.

Action 2.10: We corrected the numbering.

R2.11: Reference style is not uniform. For instance: 'Bista et al., 2015' next to 'Deiner, Fronhofer, Mächler, Walser, & Altermatt, 2016' (line 36).

Action 2.11: We had used standard APA reference style. We have updated to use the Chicago manual of style with a maximum of two authors in-text.

R2.12: Singletons consist of only one read. If the OTU consists of two reads, it is a doubleton (line 68). By the way, DADA2 is very effective in removing singletons (see Callahan et al., 2016). Thus, if you think that singleton removal '…is arbitrary and potentially hinders the detection of rare species.' you should not use DADA2.

Action 2.12: According to our understanding of the way DADA2 proceeds, the singletons removed are not arbitrarily removed but they are excluded because of low sequencing score quality (we indeed have some singleton ASVs in our datasets). Nevertheless and in order to focus on our main purpose, we removed the mention to singleton removal in the introduction, as this was not the focus of our line of thought. Rather, we present LULU as a potentially effective way to remove spurious OTUs, and as an alternative to commonly used OTU-filtering approaches, using read abundance or relative abundance thresholds.

R2.13: Though different important topics are mentioned in the introduction it is not getting absolutely clear what the authors aim to achieve and how they want to do it. Especially the late mentioning of swarm and how this algorithm will be connected to what had been said before is confusing.

Action 2.13: We have re-structured the introduction in order to underline 1) the potential necessity of clustering (at least in metazoans) and 2) the advantage of swarm v2 compared to percentage-based clustering methods.

R2.14: What do the authors mean by amplicons obtained from negative controls (lines 317-318)? They cannot possibly refer to negative controls of the PCR that yielded amplicons? I am sure there must be another explanation, but could not find it in the manuscript's methods section. There is just the cryptic sentence 'Negative extraction controls were included in each extraction run.' (line 152). Could you please explain what exactly these controls are, what you used them for and why they had been pooled with the rest of the amplicons?

Action 2.14: We have detailed the negative controls in the methods section. It now reads "*For the four field controls, the first solution of the kit was poured into the control zip-lock bag, before following the usual extraction steps. For the two negative extraction controls, a blank extraction (adding nothing to the bead tube) was performed alongside sample extractions*"

R2.15: Do more abundant species in the mock communities lead to more ASVs/OTUs?

Comment 2.15: We recovered a higher abundance of reads for those species with higher relative DNA concentration in the mock communities (see Table S6), yet, the *in vitro* admixture of DNA does not mimic the diversity of species with distinct number of cells, repeats of ribosomal sequences or mitochondria… For insects (Pinol et al. 2019 Mol. Ecol.) or fishes (Pont et al. 2018 Scientific Reports, Toshiaki et al. 2017, Mol. Eco. Ress.), correlations between biomass and read abundance could be established using very specific primer pairs. However, our global approach using "universal" primers to encompass the broadest possible range of metazoan diversity does not allow such approach. This is the reason why we prefer not to use the information about read numbers and to stick to presence/absence for metazoans.

R2.16: *Table 1: Maybe the comparison of the pipelines' results could also be presented as a figure. All these numbers separated by a slash are hard to read and may look more impressive e.g. in barplots.*

- **Action 2.16:** We have moved this table into Supplementary material and only present the read numbers for the DADA2 pipeline, as results on reads change only marginally with the other pipelines and are not the focus of the work.

R2.17: *Table 2: Could also be a 'real' coloured heatmap.*

Comment 2.17: Table 2 now includes five d-values in order to cover a broad range of clustering thresholds. It seemed to us rather difficult to visualize it as a heat map with that many d-values; we however adopted a colour code to ease the understanding of retained values.

R2.18: I struggled with the order of the paragraphs and would ask the authors to disentangle the results of the mock community approach from the results of the 'true' samples. One possibility is to restrict oneself first to the mock community results, because they allow for setting the further results in a context. Then present the alpha- and beta-diversity results of the 'true' samples.

Comment 2.18: see Comment and action 2.1.