Summary

Turba et al. (bioRxiv 2022.06.17.495388, submitted to PCI Ecology) conducted a study in turbid coastal lagoons and compared different methods for processing samples for environmental DNA (eDNA) analyses. The authors presented and tested three alternative processing methods of water samples: (1) no freezing and double filtration, (2) freezing and double filtration, and (3) centrifugation, freezing, and double filtration. Additionally, the authors tested sediment samples as a potential alternative to sampling the water column. Each of the samples were then sequenced using 12S (fish) and 16S (bacteria and archaea) primers. The authors suggest that freezing before filtering did not affect community composition for either primer, but they did find the communities targeted with 16S differed between the water column and sediment. Overall, I found this to be an interesting and important consideration of methods that can apply well beyond just the coastal lagoons sampled by Turba et al., with potential application for freshwater systems like rivers, streams, and ponds that can also be affected by turbidity (e.g., suspended particles, dissolved organic compounds). Below I offer 4 major comments and several suggestions that I hope will be helpful to the authors and their work.

Major Comments

1. Clarification about methods.
    1.1. I found the description and distinction between protocols (lines 149-155) a little difficult to follow. The description of different methods in the introduction was very well done, but I lost track of the 3 different methods to process the water samples.
    1.2. As a focal point of the paper is to show how these processing methods do (or do not) vary, it is important for the description to be abundantly clear to the reader. I think this can be achieved without too much revision, and I have provided different suggestions that might help.
        1.2.1. Suggestion 1: Have a separate sub-paragraph for each of the 3 methods, even if there are some redundancies in the text.
        1.2.2. Suggestion 2: Provide a companion table or flowchart to show the steps to process the samples across the three methods. The table could simply fill in the cells when a step (e.g., centrifugation) applies to that method, while the figure could just be a simple flow chart of arrows and text boxes to show the process for each method.
            1.2.2.1. I think even such a simple figure would be of great help to the reader and future users of the described methods.
        1.2.3. Suggestion 3: Combine suggestions 1 and 2, although this might be too much. If this approach is taken, I think the figure could be relegated to the supplement.
    1.3. Regardless of the approach taken to clarify the methods, I think it is important to explicitly state which methods applied to the no-freeze samples. I found myself

losing track of the methods and searching for how the no-freezing samples were processed in this section (lines 149-155).

2. Implement a consensus approach for differential abundances.

    2.1. Differential abundance methods can produce different results, and a consensus approach could be helpful (Nearing et al., 2022).

    2.2. Turba et al. used DESeq2 to evaluate differential abundances between protocols. DESeq2 fits a series of negative-binomial generalized linear models to count data and estimates the $\log_2$-fold change in abundance (Love et al., 2014). In contrast, the ANOVA-Like Differential Expression (ALDex) analysis accounts for community composition when calculating differential expression (Fernandes et al., 2013, 2014). These are just 2 of the possible differential abundance methods, and I am not recommending the authors run through the full list (e.g., Nearing et al., 2022), but I do think it is important to compare results from complementary differential abundance methods for a robust analysis (e.g., Fitzpatrick & Schneider, 2020).

    2.3. Turba et al. are evaluating different processing methods, but the differential abundance methods could also lead to contrasting interpretations and would therefore benefit from a consensus approach.

    2.4. Both DESeq2 and ALDex analyses can be conducted using R, and ALDex has been found to produce the most consistent and reproducible results across studies (Nearing et al., 2022).

3. Interpretation of results.

    3.1. Turba et al. provided a clear and comprehensive report of the results, but there were several occasions (e.g., lines 299-302, 304-307, 331-332, and 336-338) where more information would have been helpful.

    3.2. For example, the authors seem to rely on a hard threshold for statistical significance (alpha = 0.05), and use just the P-value to indicate evidence or absence of an effect. I highly encourage the authors to reduce the emphasis on P-values and statistical significance.

        3.2.1. I will not litigate statistical significance or discuss the broader debate on P-values in this review, but I base my comment on the following papers:

            3.2.1.1. Wasserstein and Lazar 2016, The American Statistician

            3.2.1.2. Wasserstein et al. 2019, The American Statistician

            3.2.1.3. McShane et al. 2019, The American Statistician

            3.2.1.4. Berner and Amrhein 2022, Journal of Evolutionary Biology

            3.2.1.5. Muff et al. 2022, Trends in Ecology and Evolution

            3.2.1.6. If the authors do not have access to any of these papers but would like to read them, please contact me at the electronic mail address provided at the end of the review.

3.3. P-values do not indicate evidence for an effect, as P-values just indicate how incompatible the data are with a given null hypothesis (which is already known to be false to some extent). By using P-values and hard thresholds of statistical significance, results are dichotomized (i.e., significant or non-significant). I think the authors have a lot of interesting data that should be presented in the full context, such as presenting the variance explained by a model term or contrast.

3.4. For the rarefaction curve results, slope estimates and confidence intervals would be good to report in text. These results were presented in Figure 4, but I think it might be better to have the numeric values accompany statements made in text.

3.5. PERMANOVA results were reported as either not significant (lines 331-332) or at the threshold of significance (lines 336-338); however, I do not know if this is accurate.

    3.5.1. The full R^2 for the effect of protocol in the PERMANOVA should be presented in the text or table so the readers can assess the strength of the overall effect. Moreover, the R^2 for each of the pairwise contrasts is relatively strong for both 12S and 16S (Table 2) and across rarefied and eDNA datasets for 16S (Tables 2 and 3).

    3.5.2. Additionally, after going through the rarefied and eDNA figures both in the main text and the supplement, I think there is a meaningful difference by protocol. Importantly, I do see a weak effect of freezing on community composition for rarefied 12S ($R^2 = 0.206$) and eDNA 16S ($R^2 = 0.156$) while there is a strong effect for rarefied 16S ($R^2 = 0.564$). Therefore, I think the authors may need to reconsider their statement about pre-freezing water not affecting community composition (lines 368-370).

    3.5.3. I think it would be helpful for the authors to not use a hard threshold, and instead present the results in the proper context that shows any relevant estimate, variance on the estimate, and effect size.

3.6. While I agree with the authors' use of false discovery rate correction for the differential abundance analyses, I do not think this is necessary for the contrasts. Not only are there only a few comparisons being made, but each contrast is effectively planned and not an exploratory comparison.

4. Structure of the results section.

4.1. Given the structure of the results, I found it hard to track and see how the different methods actually compared. That seems to be the main objective of the paper, but it got a little muddied in the turbid waters of the results (sorry for the truly terrible pun/joke. I know it is awful).

4.2. I like the sub-structure within the results, particularly as it started with a thorough presentation of the sequencing, bioinformatics, and data pre-processing results. However, the remaining sub-sections seemed to jump between primers (focus on

the taxa being considered) rather than focusing on the processing protocols (focus on the protocols).

4.3. The Beta Diversity section was where I was expecting to see greater contrast among the methods, but I think this will be resolved once the authors re-evaluate the interpretation of the results (particularly the PERMANOVAs).

Minor Comments

1. Lines 68-69: Could the authors clarify what is meant by "driven"? My interpretation is the organic and inorganic matter form the foundations of the food web, as driven suggests an active role, but I also know their presence in the water column is what leads to problems with filtration.

2. Lines 120-121: As I noted in my summary, I think these results can extend into rivers, streams, and ponds. No response is necessary, but I wanted to offer this to the authors.

3. Lines 190-191: I think having shared sequencing runs adds a lot of value and merit to the method. Sequencing can be very expensive and you will likely be sharing a sequencing run with at least 1 other researcher. I think this realism is a benefit to this method proposal/comparison, because it is done under completely realistic and not idealistic conditions. No response is necessary, but I wanted to offer this to the authors.

4. Lines 283-292: Great work by the authors interrogating their data and taxonomy assignments. No response is necessary, but I wanted to offer this to the authors.

5. Lines 341-342: Please correct me if I am wrong, but I was confused seeing tidewater goby presented with the 16S results when I think it should be with 12S.

6. Lines 393-394: Reference databases are one of the biggest limitations to any barcoding work, particularly in aquatic systems. I think this point might be worth further elaboration, particularly for systems that do not or can not use 12S or 16S (e.g., freshwater macroinvertebrates). I know this is not the main point of the present paper, but I think this method comparison will be useful to aquatic researchers across ecosystems and emphasizing reference database limitations is a major fact. No response or changes are necessary unless the authors choose to make revisions.

7. Lines 400-402: Very good and transparent interpretation of the methods and the inherent trades. As a single water sample was taken and this is a common method, I do not think this is a 'limitation' of the present study. I would suggest that water samples could potentially be collected from multiple points within a coastal lagoon (or any ecosystem), pooled as a single sample, and homogenized before processing. This would allow researchers to get a representative sample of the whole habitat but not increase the processing and sequencing costs (except for time and some labour).

8. Lines 409-421: Great discussion here by the authors regarding the sediment samples and the seemingly unusual lower read counts. No response is necessary, but I wanted to offer this to the authors.

9. Note on the CAP Analysis: I do not know if the CAP analysis is really needed, with a focus on the PERMANOVA and the contrasts providing the most relevant information in as simple a piece of evidence. The PERMANOVA will tell you if the protocols resulted in different community composition, with further pairwise contrasts performed to determine which individual treatments differed. Moreover, an ordination plot with either convex hulls or ellipses for each community type would be easier to interpret for the readers and quickly show if the protocols do result in different compositions.

<u>Reviewed by:</u>
David Murray-Stoker
Ph.D. Candidate
University of Toronto
dstoker92@gmail.com

Please do not hesitate to contact me directly via electronic mail if any of my comments were not clear or require further clarification during the review and revision process.

References (Peer Community Journal format from Zotero Plug-In)

Berner D, Amrhein V (2022) Why and how we should join the shift from significance testing to estimation. *Journal of Evolutionary Biology*, **35**, 777–787. https://doi.org/10.1111/jeb.14009

Fernandes AD, Macklaim JM, Linn TG, Reid G, Gloor GB (2013) ANOVA-Like Differential Expression (ALDEx) analysis for mixed population RNA-seq. *PLOS ONE*, **8**, e67019. https://doi.org/10.1371/journal.pone.0067019

Fernandes AD, Reid JN, Macklaim JM, McMurrough TA, Edgell DR, Gloor GB (2014) Unifying the analysis of high-throughput sequencing datasets: characterizing RNA-seq, 16S rRNA gene sequencing and selective growth experiments by compositional data analysis. *Microbiome*, **2**, 15. https://doi.org/10.1186/2049-2618-2-15

Fitzpatrick CR, Schneider AC (2020) Unique bacterial assembly, composition, and interactions in a parasitic plant and its host. *Journal of Experimental Botany*, **71**, 2198–2209. https://doi.org/10.1093/jxb/erz572

Love MI, Huber W, Anders S (2014) Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology*, **15**, 550. https://doi.org/10.1186/s13059-014-0550-8

McShane BB, Gal D, Gelman A, Robert C, Tackett JL (2019) Abandon statistical significance. *The American Statistician*, **73**, 235–245.

Muff S, Nilsen EB, O'Hara RB, Nater CR (2022) Rewriting results sections in the language of evidence. *Trends in Ecology and Evolution*, **37**, 203–210. https://doi.org/10.1016/j.tree.2021.10.009

Nearing JT, Douglas GM, Hayes MG, MacDonald J, Desai DK, Allward N, Jones CMA, Wright RJ, Dhanani AS, Comeau AM, Langille MGI (2022) Microbiome differential abundance methods produce different results across 38 datasets. *Nature Communications*, **13**, 342. https://doi.org/10.1038/s41467-022-28034-z

Wasserstein RL, Lazar NA (2016) The ASA statement on p-values: context, process, and purpose. *The American Statistician*, **70**, 129–133. https://doi.org/10.1080/00031305.2016.1154108

Wasserstein RL, Schirm AL, Lazar NA (2019) Moving to a world beyond " p < 0.05." *The American Statistician*, **73**, 1–19. https://doi.org/10.1080/00031305.2019.1583913