Hydrol. Earth Syst. Sci. Discuss.,
https://doi.org/10.5194/hess-2019-551-AC2, 2020

Hydrology and Earth System Sciences
Discussions
Open Access
EGU

Interactive comment on "Water tracing with environmental DNA in a high-Alpine catchment" *by* Elvira Mächler et al.

Elvira Mächler et al.

natalie.ceperley@unil.ch

Received and published: 23 January 2020

> We thank Scott Blankenship for reading the manuscript so carefully and providing constructive feedback, which will help to improve our manuscript. We addressed the suggestions and give the details of our intended improvements below each of the points. We quote hereafter the original comments (*italic*), followed by our reply (normal font) on how we plan to modify our manuscript.

General Comments

The motivation for this work was to investigate whether eDNA methods can be used to trace the flow path of water? The authors stated three questions they were pursuing.

Printer-friendly version

Discussion paper

1) Does eDNA discriminate between three sampled water sources? Two methods were directed at this question. Non-metric multidimensional scaling of species similarity. Dissimilarity measures fitted to environmental variable in what appeared to be an approach similar to a discriminant analysis or principle components.

2) How do changes in physical water flow process, reflected as hydrologic heterogeneity..., affect eDNA composition? A modelling approach was directed at this question. Species richness (i.e. rarefied of ZOTU counts) derived from eDNA given i) measure of variability in discharge or ii) an index of connectivity of water between surface and hyporheic zones. 3) Whether upstream tributaries and springs determine the eDNA signal in the main downstream channel? A comparative approach was directed at this question. ZOTUs observed from all collection locations were assigned an origin given where the ZOTU was predominantly observed. Then, the numbers of non-mainstream ZOTUs (i.e. spring or tributary) were counted from within samples taken in the mainstream to estimate the source proportion in the mainstream.

The molecular biology conducted was satisfactory to obtain species communities measures within the collections. The environmental covariates associated with eDNA collections were quite detailed. In my opinion, the field collection configuration compromised using these data to determine water path given the biological communities observed. More detailed comments are provided below, but I would encourage the authors to de-emphasize this particular study's contribution to tracing water path and emphasis the considerable amount of community ecology data available to contemplate how eDNA could inform ecosystem processes. These data are quite useful and should be published, but as a reviewer, there was an expectation mismatch between what was promised and what was delivered.

> We thank the reviewer for recognizing our work's significance for community ecology. Indeed, we consider one of the most innovative aspect of this study to be the interdisciplinary approach, allowing the use of molecular tools to help understanding hydrological aspects of understudied aquatic systems. The decision to focus on hy-

drology is based on the primary readership of HESS, an aspect that is of important for papers published in this journal (we had an exchange on this very topic with the editorial board of the journal, who encouraged us to give it that focus).

We agree a different approach would also be interesting, but our key target with this paper is to reach future “users” of eDNA in hydrology. Indeed, the study provides data for further manuscripts, which we are working on. We also plan to publish all further data collected, and look forward to it benefiting future ecological work. We will revisit the abstract and title to ensure the text and expectations created are matched with the subsequent analyses. More specifically, we will tone down usefulness as a "tracer" tool, and emphasize rather the amount of information in the eDNA, that could be the basis for a future tool development.

It was quite ambitious to suggest that researchers could peer into water collections from the base of a watershed and infer the mixture proportions (of different upstream habitat types) comprising those samples. While characterizing habitats by their species community (or even defining what an ecosystem might be) is a worthy endeavor, it is quite premature to suggest these DNA data may enable tracking a habitat’s species community through space.

> The ambitions raised in linking eDNA to different water types over multiple km can be matched with our data and is in the line of previous work. This, however, has been mostly happening within systems of one water type only. For example, several eDNA studies have shown that eDNA can travel for distances of ten to hundred km in stream systems (e.g., Deiner & Altermatt 2014, Pont et al. 2018), and it has been shown that stream networks congregate eDNA from upstream areas (Deiner et al. 2016). While the spatial scale indeed matters, the distances in our study system are actually relatively small and probably do not result in substantial decay. A recent work showed from lowland streams and waste-water inflows (Mansfeldt et al., 2020) that

mixing of different water types can be traced well and to a high level or reliability with eDNA (we will add this reference), and where this mixing could be cross-validated with hydrological information on water sources. Thus, our study fits well in this context, which we will better highlight. We added further comments below, where this point about the spatial scale was brought up in more detail.

Importantly, however, we would like to add here that we fully agree that it is premature to think that we can use eDNA to quantitatively infer water source proportions in natural streamflow as is routinely done for example with stable water isotopes (e.g., to tell how much water resulted from ground water versus ice melt). Such quantitative source tracing has indeed been proposed in the eDNA literature (e.g., Knights et al., 2011), but needs further development. Up to now it has been mostly used to infer occurrences of species under known hydrological settings, and to a lesser degree inferring hydrology from the tracking of species/eDNA. The latter can in principle be done, but indeed depends on assumptions (e.g., steady signal inflow in a given water type) that need to be corroborated, and go beyond what we do. We will carefully check the revised version to avoid any misunderstanding with respect to what we mean by “tracing”.

Rather, these data provide an empirical source of information that can be used to hypothesize how community composition could be exploited to enhance knowledge about ecosystem processes.

> This is a good point, and we will highlight it more by adding a specific paragraph. We acknowledge that the data has strong implications for community ecology and further studies from this data might go into this direction. The main intention of our study was to bridge two fields (ecology and hydrology) with the goals of: i) drawing the attention of hydrologist to the potential of eDNA and ii) to raise awareness among ecologists using eDNA of the complex hydrology which is currently ignored to a large extent. Therefore, we would like to keep this interdisciplinary aspect, which is suitable for the HESS journal, but agree that the data will allow some further analyses beyond

the scope of this manuscript.

Comments on Methods

For this review, I focused on the aspects of eDNA evaluation, and assumed that the hardware used to collect environmental covariates was operated correctly and included required quality assurance/quality control procedures. The authors used standard and acceptable lab practices. The metabarcoding primers (Geller et al. 2013) are a set currently used to evaluate the presence of metazoans, although the portion of COI targeted by Geller et al. (2013) is somewhat arbitrary. The primers appear to work well for invertebrates. The Geller reverse primer has many degenerate base pairs, allowing detection of bacteria, invertebrates, herps, mammals, etc. The authors' bioinformatics procedures look typical of approaches other labs have taken to metabarcoding. They only instance where the authors made a procedural decision not cited in published literature was the choice of removing any sample with sequence below 20,000 reads. They declared them "failed". This seems reasonable (2.4.2 Data Cleaning, last line in paragraph).

> Thank you for your competent and profound evaluation of our methodological descriptions, we agree with your summary. We appreciate that you reviewed the methodological aspects to verify this integral part of our paper. This is especially important for the interdisciplinary audience, potentially unfamiliar with such methods. Indeed, the COI primers by Leray/Geller were originally developed to identify stomach contents of fishes, but are currently the most popular primers for eukaryotic work. We performed rarefaction to the sequencing depth of the lowest sample; therefore, we removed the sites with suspiciously low read numbers to not further constrain the data too much.

Comments on Results

1) Does eDNA discriminate between three sampled water sources? Differences in species composition were observed by habitat type. This is useful information. Yet, it

is unclear whether the differences were divergent enough to support mixture analysis.

> For now, eDNA does not discriminate completely between the three water types. While such a discrimination would be ultimately wanted, our dataset alone would not yet allow to describe it. We agree that there will likely always be a portion of the eDNA signal that is non-informative on the source types, yet the goal is to identify "marker ZOTUs", that is signals that are specific to one of these water types. The feasibility of this has been indicated by a mixture analysis for eDNA in a recent study testing the influence of waste water treatment plants on pristine streams performed in a highly replicated manner (Mansfeldt et al. 2020). The authors used a Bayesian approach to estimate the proportion of contaminants in a given community that come from possible source environments, based on the identity of sequences (Knights et al. 2011). In a context of completely distinct sources, a source tracking is possible. In systems as the one studied, with variable and overlapping signals from different sources, a mixture analysis is not yet possible (and we also didn't want to imply this). We will highlight this better, and discuss it in the context of the recent study by Mansfeldt et al. (2020).

Perhaps the following two examples will prove useful, I don't know. They relate to how I was intellectualizing tracing water. Mixture analysis is a common population genetics application. Procedure is predicated on sources contributing to an unknown mixture being statistically distinctive (at population level of allele frequencies). Distinction enables a probability to be derived for each observation, with mean likelihood over all observations essentially being the estimated mixture proportion. A second example is decoding movement patterns using otolith microchemistry. If an individual moves through space and time being exposed to different isotopic ratios, these differences are reflected in the otolith. Again, this is predicated on differences in chemistry across space. In the case here, what is the probability a group of species may be observed in a habitat type? If species (probabilistically) could occur in many or all habitats, then the power to identify a habitat using species is quite ambiguous. There are additional comments below in #3.

> These are interesting examples. In hydrology, this kind of unmixing approach is in fact common (e.g., Beria et al., 2019, with an application to the same catchment); such studies either use differences in water chemistry across space (to trace how much different geographic source areas contribute to streamflow) or differences in time (e.g., to trace how much snowfall versus rainfall contributes to streamflow). While "temporal" tracing is relatively straight forward with the help of stable isotopes of water (in precipitation, they always show some seasonality), geographic tracing remains difficult except in places with very particular geological (i.e., chemical) settings. The omnipresent characteristic of eDNA is exactly why tracing with it could become an interesting tool in the future. However, eDNA as a tracer has two main challenges: eDNA is not a conservative tracer (i.e., the concentration changes in time in the receiving water body and is also variable in the source) and not all sources of eDNA can be sampled or are known. These challenges are classical in hydrological tracing and can be addressed in future work. For an example how to model decay see the work of Carraro et al. (2018).

Exploring how eDNA could actually be useful in such a natural (alpine) setting is exactly the scope of our paper. We will make this clearer in the introduction by giving more context on source tracing in hydrology but also in genetics.

We would like to note additionally that while quantitative water source tracing with eDNA is yet to come, the recent work of Carraro et al. (2018) proposes a hydrological modelling framework to quantify potential source areas of eDNA of specific species. The study uses an opposite of our approach compared to us. They assume that water flow paths and quantity are known but would like to quantify the eDNA sources areas. This is obtained via a spatially-explicit model of how eDNA is transported along the river network, combined with a local eDNA production model and assumes constant hydrologic transport conditions in time (mean streamflow) and only one hydrological system (river network), and does not separate different hydrological sources, such as ground- or melt-water. The model is calibrated on observed eDNA. The framework of Carraro et al. corresponds to a classical inverse modelling approach, one where we

know the target, in this case eDNA in the streamflow, and obtain a solution to model this target by model calibration.

This is in opposition to source unmixing approaches, which correspond to forward modelling: e.g. one knows the source composition and tries to quantify how much these sources contribute to the target. We will include these explanations in the revised version and refer to this response, since they shed further light on the question of how to use eDNA in hydrological studies and vice versa.

2) How do changes in physical water flow process, reflected as hydrologic heterogeneity..., affect eDNA composition? I understand the interest in species composition. Further, the data provide by the authors provides direct evidence that both habitat complexity and connective will increase species richness. Yet, the number of species isn't indicative of a habitat type. Perhaps the exact timing and location of a subset of species could predict habitat type, but these data are not presented.

> Indeed, richness is a relatively simple measure, but it is often used as a basic investigation for diversity analysis and thus is a good and broadly accepted starting point for a first assessment of eDNA in an alpine catchment. We also adjusted the term 'eDNA composition' and changed it to 'ZOTU richness', which is actually the measure we use in this part. We agree that further investigations are needed, specifically including potential shifts in identity. We therefore compared the eDNA samples among the three stream types (original manuscript Fig. 4, NMDS). This analysis is based on species identity (i.e., the dissimilarity of species at the individual sites) and not on richness. We ran a preliminary test to identify candidate ZOTUs (or organisms when taxonomic assignment is available). This allows construction of lists of species that are associated with particular groups of sites (or combinations of those, De Caceres & Jansen, 2016). We found 418 ZOTUs significantly associated with the main channel, 185 to springs and 1369 to tributaries. 324 ZOTUs were associated with tributary and main channel, 7 between springs and the main channel, and 33 between springs and tributaries. We are happy to include this analysis in an updated version of this manuscript.

3) Whether upstream tributaries and springs determine the eDNA signal in the main downstream channel? In this question is where it all went wrong for me regarding using eDNA to deduce water path. The species detection distance is undetermined. Pronouncing a sample as "spring" because a species was observed in spring and mainstream, but not tributary, is dubious without accounting for distance (effect size) of DNA detection for metabarcoding given your lab's procedures. The assumption is that each species has a 100% detection probability below its origination point, which isn't the case.

> We agree that the spatial scale matters, however the distances in our study system are relatively small and can be assumed to have minor influence. Thus, the likelihood of decay during these transport distances can be neglected as a first approximation (in the literature, distances for substantial decay are in the order of 10 to 100 km). We add now a table to the manuscript (Table R1 in the addendum to the reply to the first reviewer), indicating the distances of each sampling site to the outlet and the distance to the main channel for springs and tributaries. The maximum distance of a point to the outlet is 4.2 km and the maximum distance of a tributary/spring to the main channel is 100 m. Several eDNA studies were showing, that eDNA can travel well beyond this range (e.g., Deiner & Altermatt 2014, Pont et al. 2018), 10 to 100 km. Detection probability, however, is indeed a critical point and is yet largely neglected in metabarcoding studies. The detection and assignment of individual ZOTUs to these their water types indeed depends on abundance, distributions and a specific detection probability. This is indeed not easy to disentangled, and it is true that the absence of a signal in a tributary does not a priori indicate that it is an exclusive spring signal (because there can be also false absences, or absences due to low occurrence and thus the likelihood being low). However, ZOTUs having a high detectability between habitats would blur the pattern and be against our hypothesis, thus all findings of ZOTUs being distinct for habitats are indeed suggestive of this pattern being real. In Fig. 6A, we see indeed that there is a large overlap of the three sites (i.e., the center of the Venn diagram), indicating that we likely underestimate the contribution of tributaries and springs. The focus is, however,

not on those but rather on the ZOTUs that are distinct, and there is a sufficient number of ZOTUs that are not shared.

Habitat types are aggregated. Yet, each site is spatial explicit. For example, were all species from tributary UT observed at River UR, with a predictable reduction at BR, then HR, and subsequently MR/ER. Another example, spring GS is below every collection site expect River MR/ER, spring HS is below half of collection sites and spring BS is below two tributary and two river sites. In my view, aggregating data (i.e. species) by habitat type ignores that geographic location of sites precludes the movement of species across habitat types for many (most) study sites.

> We apologize if the scale of the map was not so clear, and some of these spatial layouts not visible. We will include a new table R1 in the main document or supplementary file to clarify. It is important to note that upstream sites only contribute to downstream sites of the main channel, but not to any other site type (i.e., neither to other "spring" nor "tributary" types). The catchment is very steep, and all other sampling sites are a substantial distance from the main channel away, so that "upstream" mixing is impossible. In that sense, it is not possible that other sampling site (besides the main channel) receive water from upstream sampling sites. Of course, organisms may move between sites in other ways than the downstream flow, but in this case, absolute distance would be a better metric than the channel network. We have included a new table that shows the approximate distance from each tributary and spring sampling site to the main channel, and will clarify in the manuscript the distinctness of the tributary and spring sites, and state that mixing could only occur in the main channel. The main channel itself by definition captures all different sources, as this is inherently found in such hydrological systems.

To trace water path, species from a habitat type (source) would have to be absent in receiving habitat type, with detection of that species in receiving habitat being indicative of source contribution. Saying there was a shift in species richness is inadequate. Perhaps specific species were diagnostic for richness changes from different sources?

This was not presented, but if it exists could form the basis for a subsequent project developing a water path analysis.

> We apologize for a possible misunderstanding, and will clarify. As explained above, indeed we cannot distinguish between ZOTUs that were shared by the different water types vs. ZOTUs that were just transported downstream, but absent from the receiving main channel; however, the latter can only be happening in the main channel, while distinctness in tributaries and springs can be established. We run a preliminary test to identify candidate ZOTUs, allows determining lists of species that are associated to particular groups of sites (or combinations of those, results shown above).

Further, was there a seasonal shift in species detected? ZOTUs are a fairly liberal index of diversity. For species noted by name (i.e. unverified species of midge, short-horned grasshopper, alpine salamander, common European adder, ptarmigan, and the chamois) were each of these species present throughout entire study? (Could be any species though).

>In the tributary and main channel sites, there were pronounced seasonal changes in the number and identity of ZOTUs, while in the spring sites the signal was much more constant. This is consistent with the different hydrological characteristics (one being fed by groundwater, while the other being dependent on precipitation and snowmelt). We agree that ZOTUs are a relatively liberal index, unfortunately, however, it is currently not possible to assign species names to most of these ZOTUs: in central Europe, for most of the groups expected (invertebrates, especially), the database coverage is less than 50%, and for such specific high alpine areas it may fall even below 10%. Consequently, we were only able to put a taxonomic name to a small proportion of our ZOTUs, resulting in 189 different species names; those did not show pronounced seasonality. Nevertheless, you are correct our metric could have some seasonality signal of species that might co-concur with the seasonality of hydrologic processes, and could thus disguising the physical processes that we want to detect. We will add a few sentences clarifying this, and refer to a key study on seasonality signals detected

using eDNA (Bista et al. 2017).

Perhaps natural introduction and removal of eDNA source at (upstream) habitat types could be informative for water contribution from that site downstream to other sites connected by water flow. Again, if species are present throughout different habitat types and elevations, then using those species to identify water source is limited. > This is an interesting point, and we agree that different upstream habitat sources can (and will) influence a downstream signal (see also Mansfeldt et al. 2020). However, such complexity of mixing is true for most natural occurring tracers, including stable isotopes, conductivity, various solutes, etc. An approach to use DNA as a tracer would be to identify a single species that is uniquely present at each sampling site. Then this species' DNA could be studied intensively in terms of quantity, travel distance, etc. and serve as a downstream monitoring tool. Such a tool, however, would need to be developed on a site-by-site basis and would require intensive pilot studies. It would also not allow to be transferred to other areas and would be more exploratory and less universal than our approach.

Data from this manuscript provides a means to explore these issues further. This is a beginning step in this endeavor and would benefit from adjusting the tone of the manuscript to reflect the infancy of this line of research. Toward the end of the discussion the authors note the study design did not capture the complexities involved with the overarching objective. While I agree, I would also stress that the spatially explicit information about species diversity will prove useful for the authors and other researchers to further explore ecosystem processes.

Thank you again for your review. We agree that the field as such is in an early stage, and we will adjust the formulation of our statements, and be more cautious. We will emphasize the significance of our study for ecology and be more upfront about what we expect to find with our metabarcoding approach using eDNA, and where we see the possible limitations. We also will publish the data along with this paper, which provides a concrete base line for future ecological work. We agree with you that our

study can be a starting step for many other studies to come. This is also why we would see our study a good fit for HESS, because it is a conceptual but also quantitatively corroborated proof of a new concept, merging tools from ecology with hydrology, to get a better understanding of complex aquatic ecosystems.

Finally, we would like to make a general comment regarding the use (and value) of ZOTUs vs. species identity. The reviewer correctly states that using ZOTUs does not tell us about the identity of the organisms behind. We have already mentioned that the lack and incompleteness of the databases does currently not allow us to do such a specific assignment. We agree that for certain questions, such as in conservation biology, the identity of the species would be useful information, and we also used this information for the subset of ZOTUs that had a matching entry in the database to verify that the organisms observed are actually realistic to be found in these alpine areas, which is the case. We do, however, also want to mention that for our question, namely studying how eDNA can be used to track and differentiate qualitatively different water types, it is actually not needed at all to have the species assignment done. It is sufficient to do the analysis on the ZOTU level, as all we are interested is in the shared vs. unique signal from different water types. This could be species, ZOTUs, or even a different signal. The advantage of the ZOTU level is that we do not restrict ourselves to a subset of organisms that are commonly used in aquatic ecology, or those that are covered in the databases, but are using an overall "fingerprint" of organisms. Thus, biologically/ecologically one could gain more information when knowing the taxonomic identity of the ZOTUs, while for our study question on hydrology, the ZOTUs are actually perfectly fine, and we might have done the same analysis even if there were complete databases. We plan to add this information in the manuscript, or are happy to refer the interested reader to this response.

References

Printer-friendly version

Discussion paper

Beria, H., Larsen, J.R., Michelon, A., Ceperley, N., & Schaefli, B. Hydromix v1.0: a new bayesian mixing framework for attributing uncertain hydrological sources. Geoscientific Model Development Discussions, 2019.

Bista, I., Carvalho, G. R., Walsh, K., Seymour, M., Hajibabaei, M., Lallias, D., ... & Creer, S. (2017). Annual time-series analysis of aqueous eDNA reveals ecologically relevant dynamics of lake ecosystem biodiversity. Nature communications, 8(1), 1-11.

Carraro, L., Hartikainen, H., Jokela, J., Bertuzzo, E. & Rinaldo., A. Estimating species distribution and abundance in river networks using environmental DNA. Proceedings of the National Academy of Sciences, 115(46):11724–11729, 2018.

De Caceres, M., & Jansen, F. Relationship between species and groups of sites. Package 'indicspecies', version 1.7. 6, 2016.

Deiner, K., & Altermatt, F. Transport distance of invertebrate environmental DNA in a natural river. PloS one, 9(2), e88786, 2014.

Deiner, K., Fronhofer, E. A., Mächler, E., Walser, J. C., & Altermatt, F. Environmental DNA reveals that rivers are conveyer belts of biodiversity information. Nature communications, 7, 12544, 2016.

Knights, D., Kuczynski, J., Charlson, E. S., Zaneveld, J., Mozer, M. C., Collman, R. G., ... & Kelley, S. T. Bayesian community-wide culture-independent microbial source tracking. Nature methods, 8(9), 761, 2011.

Mansfeldt, C., Deiner, K., Mächler, E., Fenner, K., Eggen, R. I., Stamm, C., ... & Altermatt, F. Microbial community shifts in streams receiving treated wastewater effluent. Science of the Total Environment, 709, 135727, 2020.

Pont, D., Rocle, M., Valentini, A., Civade, R., Jean, P., Maire, A., ... & Dejean, T. Environmental DNA reveals quantitative patterns of fish biodiversity in large rivers despite its downstream transportation. Scientific reports, 8(1), 10361, 2018.

Interactive comment on Hydrol. Earth Syst. Sci. Discuss., https://doi.org/10.5194/hess-2019-551, 2019.

Printer-friendly version

Discussion paper