02 Aug 2023

Dear Editors and reviewers.

We are happy to have received a response concerning the revised version of the manuscript that we submitted in March 2023. Of course, we are interested in further improving our manuscript as indicated in our responses below. However, we would like to note that we are rather surprised by the protocol. Firstly, there is no mention of the adequacy of our responses to the 2 reviewers initially assigned to this manuscript. You have indicated that the manuscript is still in need of minor revisions, yet there is a full-blown second round of review suggesting to us that we were very inadequate in our responses to the initial reviewers. However, the new reviewer raises questions that were not covered in the first review, making us wonder how the first round of review was useful. We also wonder if this new reviewer is not available to follow-up on our responses to the criticisms, will we need yet another round of full review? Based on our experience in publishing manuscripts, it is usual that the editor judges if the responses are adequate if the initial reviewers are not available. It would be useful to announce the specific policy of PCI to potential authors as it is not habitual.

Here we indicate, point by point, how we have responded to the reviewer remarks.

Our responses to the reviewers' remarks are in **blue bold face** below each remark. Text highlighted in grey is part of the message received from the editors that did not require a response.

Thank you again for your efforts.

For all of the authors,

Cindy Morris

= = = = = = =

Round #2

Author's Reply:

---

by Mina Bizic, 11 Jul 2023 12:01
Manuscript: **https://www.biorxiv.org/content/10.1101/2022.09.06.506731v2** version 2

Further minor revisions needed

Dear authors,

I am sorry for the time it took to evaluate your revised preprint. The original reviewers were unavailable at this time and it took a while to find new reviewers.

Your revised preprint has been now reviewed by one peer who was made aware of the previous comments and revisions. As you can see this reviewer, finds your work very interesting and the manuscript well-written, and has several more suggestions and comments for improvement.

Most of the comments are minor and can be rapidly addressed.

> **Please note that many of the comments suggest re-analyzing data and re-structuring the logic of the presentation of information. This involves lots more than a minor investment in time. In light of the effort by this additional reviewer to make numerous comments, we do not feel justified to treat them as minor remarks.**

With regards to the suggestion to remove the term "seasonality" from the title. Indeed the main focus of the paper is not the seasonality aspect. Though you did sample all seasons across two years, not much emphasis is given to the temporal aspect. Therefore, I suggest considering accepting the suggested title. Alternatively, as also suggested, please enhance the temporal aspect of the discussion and analysis.

> **We have removed "seasonality" from the title.**

I am looking forward to your revised version,

Mina

Reviews

Reviewed by anonymous reviewer, 03 Jul 2023 13:41

Morris et al. conducted a study about the abundance and diversity of two plant pathogens *P. syringae* and the pseudomonadal family Pectobacteriaceae in the river catchments in France

> **We note that the Pectobacteriaceae are in the gamma Proteobacteria as is *P. syringae*, but that they are members of the order Enterobacterales and can be facultative anaerobes unlike *P. syringae*.**

Title/abstract

The manuscript is well written and thoroughly structured. The description of the rationale is clear from the title and the abstract. The abstract comprehensively summarizes the results from the study including the abundance of both plant pathogens of interest and their assessed diversity. Furthermore, authors give an outlook on the discussion in their manuscript. I'd suggest two things regarding the title and the abstract: 1) maybe it is not necessary to call the study "comparative abundance and diversity" because in my view it is not clear why these two plant pathogen (groups) stand in comparison/contrast to each other? Do they? If yes, it should be stated to understand the importance of the comparative character of the study.

We modified sentences near line #48 to clarify the differences. See responses to "Introduction", below.

Additionally, as authors try to cover the population abundances from the entire year, seasonality is not a focus of the study and might be left out from the title as the reader somehow expects the seasonal effect to be presented in more detail. Thus, the Title could be: "Abundance and Biodiversity Patterns of Pseudomonas syringae and Soft Rot Pectobacteriaceae Species Complexes throughout the Durance River Catchment: From French Alps Sources to the delta". If the authors would like to put more focus on the seasonality, I'd expect more analysis, figures.

We have modified the title by removing the word "seasonality". This change avoids making a 2-part title with a colon as punctuation.

Furthermore, I think the two sampling years 2016 and 2017 could be combined. Please test the difference of the years and consider putting the results together.

For both organisms, bulking the 2 years would mask the variability that we wanted to reveal and that are rather contrasted for the 2 groups of bacteria. See the sites R01, R02, T02 and T03 for SRP as examples of the information that would disappear by doing this. Furthermore, for analyses concerning correlations with water conditions, we have the full scope of data for water conditions for 2017 only. Hence, the data need to be separated by years for this purpose.

2) The Abstract ends with an outlook on the discussion. I'd prefer to have 1-2 sentences in the end that summarize this outlook, meaning that summarizes the discussion and the ideas of the authors of how to incorporate this knowledge in a strategy for anticipating risk for disease outcrops in a catchment.

Thank you for this suggestion. The end of the abstract has been modified.

Introduction

The Introduction reads quite well and is interesting. I'd like to make 3 suggestions for the introduction. 1) As said in the paragraph above, in my view it might be good to emphasize what is missing in the current literature and which gap the authors would like to fill with their study.

The state of knowledge and the deficit of information for plant pathogens is described in the first 2 paragraphs of the introduction.

2) Furthermore, I'd suggest to put the research question/hypothesis at the very end of the Introduction and moving the explanation of both plant pathogens before the last paragraph.

We feel that this change would alter the logic of our presentation. This work is about the scientific questions in general concerning the surveillance of plant pathogens in river water. The biological models are not the bases of the central questions.

Also clearly formulate 2-3 research questions of the study also including the target to explain the variance in the abundance and diversity which the authors emphasize later,

We have added a sentence in the 4th paragraph to clarify this.

3) Here in the Introduction again, I think it would be good to clarify in detail why you put focus on the comparative character of the study. The results later do not really emphasize this. If yes, it was not very obvious to me.

The reasons for comparing these 2 groups of bacteria are described fully in the last paragraph of the introduction. We have modified sentences in this paragraph to clarify this.

Materials and Methods

The method section is very detailed (which is good), still I can sometimes not follow the workflow that is described (pls see below). The measurement of the chemical properties of the water (DOC, NH4, etc) is only explained by citing a reference. I understand that it might be well explained in the cited literature but I think it would be good to shortly explain the method you used to quantify the properties. CFA? Color reaction? This allows the replication by others researchers more easily. Although these are the describing metadata, the results are always related to the abundance and diversity of the plant pathogens and play a prominent role in your manuscript.

We have added more details about the chemical analyses used in this work.

Apart from that, I think that some explanations in the manuscript could be shortened and be made more concise in terms of grammar/flow of the text. In terms of statistics, I have some concerns regarding the main factors from the PCA (see Results chapter).

We responded below to the questions about PCA analysis. Concerning grammar and word flow, we modified where the reviewer pointed this our specifically.

Some details for the M&M section:

In line 110, repeat the number of replicates and the volumes again.

OK, we added this.

In line 138: maybe explain bona fide.

We added the phrase, "according to previously established criteria".

Line 141: this is not clear. So they were intentionally placed twice? How did you decide on the subsample that are run twice?

Yes, they were placed intentionally to verify reproducibility as explained in the text. The strains were selected randomly; this precision has been added.

Line 149+150: please give the details for the brackets (see below, described below) here or refer to in the section below more clearly. I cannot exactly understand the usage of the 142 wells.

In this section we state very clearly that we used three types of negative controls: 1) ultrapure water (80 wells), 2) ultrapure water + PCR mix (414 wells) and 3) black wells (142 wells). We have modified to make this clearer and to better indicate the meaning of the parenthetical phrase.

Line 152: bp

Thank you for noticing this

Line 153-154: for me it is not totally clear if you used all fw primers or not.

The sentences have been modified to indicate how all of the primers were used.

I would merge the chapter "Identification of Amplicon Sequence Variants, ASVs" with the chapter above. I don't think it is important to have this separated.

OK we merged, but also lengthened the heading of the section

Line 193ff: unfortunately, this is also not clear to me. Here you refer to copy numbers, but do you mean read numbers?

We think that the reviewer is referring to lines 196-199 because "copy number" is not in line 193. Where we use the term "copy number" were are referring to the number of ASVs which is the product of forward and backwards reads. Hence, "copy number" is the correct term. The sentence in lines 183-185 was modified to explain this. We also made modifications in the section "Verification of controls, filters of ASV, and analysis of replicates" to further specify why we say "copy number".

Line 218-225: please revise this paragraph, it does unfortunately not read very well and could be shortened/be written more concisely I think.

We made a few modifications to help clarify the tools used for statistical analyses.

Line 218-225: did you use Spearman rank correlation because the data were not normally distributed or no linear connections could be expected. Maybe better explain that here.

Spearman's rank correlation is a non parametric test that avoids assumptions about normality and equivalence of variance that we cannot necessarily validate. This is a robust, widely used test.

Line 219: delete "with modules"

OK

Line 222+223: This statistical package was also used – delete this and write: Statistica was used to calculate regression…..

OK

Results + Tables and figures

Throughout the manuscript, better replace altitude with elevation

This is a major comment. When collecting data on location of our sites, the word "altitude" is proposed by many data sources. Google Earth and the US National Oceanographic and Atmospheric Administration (NOAA) website use "altitude", for example. We are not opposed to the word "elevation" as it may indeed be the more precise word. However, in the data that we have deposited on https://entrepot.recherche.data.gouv.fr we used "altitude". This is described in our metadata section of the submission. Therefore, if we changed to "elevation" in the manuscript we would be inconsistent with the deposited data. Is the reviewer suggesting that we also change all of our deposited data for a modification that does not influence the understanding of the variable we are reporting?

Table 1: please add more information to the Table headline

Could the reviewer please specify what is missing here? This table indicates the sites where we sampled in the Durance catchment – as stated in the title of the table and as described by the column headings.

Fig. 1: better write log10 bacteria L-1 in the y-axes label

Thank you for catching this. The image was a bit chopped when we pasted it in.

Fig. 1: as written above, may consider to sum the two consecutive sampling years.

As indicated above, for both organisms, bulking the 2 years would mask the variability that we wanted to reveal and that are rather contrasted for the 2 groups of bacteria. See the sites R01, R02, T02 and T03 for SRP as examples of the information that would disappear by doing this.

Fig. 2: please add information on the boxes (median, mean) in the Figure text.

This has been added to the figure heading.

Line 244: geographic situation sounds strange.

This has been changed to geographic location.

Fig. 2: please include only significant results, so better remove the pink background because I think you should only include significant results and label them with the background as you did. To include p-values up to 0.1 for significance is not acceptable.

The p-value is a measure of the probability of the differences in means from observations of two populations being due to random error rather than the "treatment" to which each population was exposed (or not). The rules for use of a cut-off for p-values to be considered as a significant effect is a matter of convention. 5% is

considered to be an "acceptable" level of risk; but why not 1% or why not 6%? When the probabilities are reported then the reader can make an informed decision. Here we are presenting the trends and the associated probabilities hence the reader has all the information to make their judgement. In our opinion, this is much more informative that ignoring the trend for Psy relative to water temperature, for example. The reader can appreciate that there is a trend of negative correlation of Psy populations with temperature throughout the catchment but that this trend is less marked in the lower basin.

Line 340ff and Table 3:

Do you mean you chose 7 main/composite factors from the PCA? According to my experience, this is way too much. Normally you end up with 2-3 main components from the PCA that still add explanation and then all the additional components do not explain much further. So, I'd encourage the authors to go through these data again and concentrate of 2-3 main components. A further question comes up: when looking at your datasheet of all the water characteristics: how did you account for the incompleteness of the dataset? Phosphate and N-components were measured in 2017, but not in 2016, meaning you have 3 water characteristics in 2016 and 7 in 2014. Does it say in the manuscript? I cannot find the paragraph where you describe that you measured these parameters in 2017 only and also not how you accounted for this imbalance of dataset when doing the PCA. Please explain.

Regarding the PCA. In my view it does not make sense to have 7 factors when you include 7 properties in the model. The PCA has the idea to reduce the variables, by calculating main components. If you take all 7, then you could also take all 7 variables as they are. For that, I recalculated the PCA with the data from 2017 only. I would reduce the PCAs to 3 main components, because they already explain ~70% of the variance and are significantly a function of temperature (PC1), DOC (PC1) and NH4 (PC2) and PO4 (PC3) (I maximized the variance) which perfectly emphasize and confirm your results.

Furthermore, the statistics you continue to do with the main components are not necessarily needed and don't add further value/benefit to the manuscript. Apart from the suspicious 7 main components, the data in Table 3A and correlation in Table 3B are just a function from themselves. I would recalculate the main components and definitely leave out Table 3.

As stated on line 340, the PCA analysis was conducted for the 79 observations in 2017 – the year where all variables were assessed.

We thank the author for confirming that reducing the number of PCA factors leads to the same results as we obtained. Our intent with the PCA analysis was to use the subsequent multiple regression step to identify objectively, transparently and in a step-wise manner the water variables that contributed to explaining the variability of bacterial populations.

We identified the water variables that explained 80% of the variability in population size. We have now modified Table 3 to indicate the contribution of each factor to the overall variability of the water conditions. The first 4 factors explain about 75% of the variability and 5 of the seven factors are needed to explain at least 80% of the variability in the water conditions (we note that the reviewer proposed to reduce the number of factors to explain 70% of the variability). We have added remarks in the text to note that there are few correlations between water variables; hence, it is understandable that PCA does not help a lot in reducing the number of factors to consider. This analysis then allowed us to rank the contribution of each factor and to assess which combinations consistently contributed to explaining the variability for each group of bacteria (total, Psy and SRP). We note that *F*6 contributes significantly to explaining the variability in SRP population sizes although it is outside of the *F* factors that explain 80% of water condition variability.

As the reviewer notes, we obtained the same results as via their method. However, we feel that our method is totally transparent and step-wise in terms of how we identified which of the physico-chemical properties of water contributed to bacterial population variability.

Line 420: MLST is mentioned for the first time. Please explain in M&M

This is a description of reported results (not ours). We have added the definition to the text.

Figure 3: is fine and show the main result of the manuscript and goes in line with the findings reported in Figure 2 already.

Discussion

The discussion is well written and the data are not over interpreted. A big part of the discussion raises questions and gives ideas of possible integration of the data in risk management studies, but in my view these ideas should be more concise (Lines 655-663 are nice but a little bit vague) if they should be part of the manuscript and the study.

The reviewer remarked that we should be more precise in the abstract about how our results can be used in risk management. We have modified the text to help clarify the points we are making about surveillance.

Line 527-528: can you report more about the land-use in the 3 different basements and the respective elevations.

We have added details.

Line 664-665: this is not really done in the study.

The authors have contributed to the question about where are the reservoirs of plant pathogens, particularly for *P. syringae*. These works are cited in this manuscript. This current work is the continued pursuit of how sampling substrate influences the picture we build of the diversity of the metapopulation of plant pathogens and how samples from agricultural fields can limit the scope of diversity that is perceived.

Maybe authors should discuss the results in the context of the raised (or to be raised) research questions.

Is the reviewer suggesting that we make the discussion even longer? The second to last sentence in the discussion (about being able to know which pathogenic lines are actually thriving in the environment and which are just in transit) sounds like a major research question.