12 November 2015

Dr. Isabelle Laurion,
Associate Editor - Biogeosciences,
European Geophysical Union

Dear Dr. Laurion,

RE: Co-occurrence patterns in aquatic bacterial communities across changing permafrost landscapes by J. Comte, C. Lovejoy , S. Crevecoeur, and W.F. Vincent

Thank you very much for your consideration. Please find attached our revised manuscript with changes highlighted in red.

We are grateful for the detailed comments made by the reviewers, which have strengthened the manuscript. All suggestions have now been incorporated in this revision, and our detailed responses to comments by the reviewers are given below. Line numbers in the responses refer to the submitted revised manuscript.

We thank the editor and the reviewers for their insightful contributions and interest in our work. We reconfirm that this material has not been published or submitted for publication elsewhere, the content and authorship of the submitted manuscript has been approved by the co-authors and there is no conflict of interest.

Sincerely,

Jérôme Comte, PhD
Département de biologie - Université Laval
1030, avenue de la Médecine
Québec (Québec) Canada
G1V 0A6
Email: jerome.comte@takuvik.ulaval.ca

CC all co-authors

Anonymous Referee #1
I was very pleased to read the manuscript "Co-occurrence patterns among aquatic bacterial communities across changing permafrost landscape" by J. Comte et al. The paper is well written and it takes an advantage of the statistical tools that are available for this type of data. The paper describes similarities and differences in microbial communities residing in thaw ponds within and between 5 regions in the subarctic Canada, along permafrost gradient. It concludes that these communities are heterogeneous, just as these environments are, but still share many key taxa. Further, they suggest that bacterial communities within a valley are more similar at the areas of

extensive melting of the permafrost and that the properties of the bacterial network make these communities robust against environmental change. Authors have done an extensive statistical analysis on the community data.

We thank the reviewer for this positive feedback on our manuscript.

What I was missing was the environmental data. It is stated multiple times that especially DOC and conductivity had a big impact, but would be good for the reader also to see how big differences there actually were among and within valleys.

We have now added SI Table 2 (former SI Table 2 is now SI Table 3) in supplementary information. This new table presents data for the environmental variables measured in the ponds.

Also please include to the table column indicating how the valleys are situated in relation to permafrost degradation gradient.

Following the reviewer's suggestion, we added a column indicating the type of permafrost landscape that each valley was located in.

Another issue that should be taken into account through out the ms is that usually classifications of 16S genes, especially for such a short fragments as here, only suggest that the OTU is closely related to certain organism. Thus, authors should be very careful in making conclusions about the metabolic capacity of an OTU based on 16S.

We thank the reviewer for the comment. This is an important point and we have now rephrased the text where any metabolic inference of OTUs was made. For example, on page 19, lines 552-556, the text now reads: "OTUs closely related to bacteria in the Chitinophagaceae, a group known to be involved in the degradation of chitin and other complex polymeric organic matter (del Rio et al., 2010), were well represented in our study area, and have also been found in other cold terrestrial environments (Franzetti et al., 2013; Ganzert et al., 2011)".

Specific comments
L65 "Specifically, patterns of co-occurrence may reveal to what extent groups of microbes share habitat preferences, to what extent there may be ecological linkages among bacterial taxa and with other planktonic organisms, and the extent of phylogenetic closeness of co-occurring bacterial taxa." I'm not sure if the patterns of cooccurrence per se reveal the extent of phylogenetic closeness of co-occurring bacterial taxa. Please explain.

We agree with the reviewer that co-occurrence patterns do not provide information on phylogenetic closeness. The hypothesis behind the statement was that phylogenetically

related OTUs may tend to co-occur more than individual OTUs given that they may share life strategies and traits. We rephrased this part as follows: Page 3, lines 71-72: "[…] and the extent of phylogenetic closeness of co-occurring bacterial taxa where closely related taxa may co-occur given that they may share life strategies and ecological traits".

L271 I have problems understanding this sentence. Do the authors mean that communities within SAS were similar and different from communities from other valleys, or that within SAS communities were more similar than within other valleys? I'm assuming it is meant to be the latter option, how ever; I'm not fully convinced that SAS and KWK had more similar communities than others. I get the impression that this conclusion is based on visual inspection of the figures. Was, for example, average distance to centroid shorter for those valleys than for others?

Yes, our aim was to investigate whether communities within southern valleys (SAS and KWK) were more similar phylogenetically than with other valleys. In response to this comment, we have performed the analysis suggested by the reviewer and results from betadispersion analysis showed no difference in the distance to centroid among the different valleys such that no bacterial communities within a valley appeared to be more similar (shorter distance to group centroid) or variable (greater distance to centroid) than among the others. We have thus rephrased this section that now reads (lines 289-298): "The clustering and principal coordinate analyses (PCoA) based on weighted UniFrac distances (dw4000; Fig. 1A, 1B) suggested that communities within SAS tend to cluster together (more similar) such that within SAS valley, communities were more similar than within other valleys. The same pattern seemed to occur in KWK valley. However, a test for homogeneity of multivariate dispersions did not support this as no significant difference in the distance to group (valley) centroid was detected (P=0.39, F=1.08). Permafrost landscape type had a significant effect on phylogenetic composition (Permutational analysis of variance on dw4000; R^2= 0.31, P=0.001). The reference lakes did not group together, likely reflecting their disparate catchment properties."

L279 This sounds like the result was depending on the location in the analysis; please rewrite.

We rephrased as: "Cluster analysis based on unweighted UniFrac distances indicated a stronger clustering according to permafrost landscape type (Permutational analysis of variance on duw4000; R^2=0.51; P=0.001) by comparison with weighted UniFrac distances (SI Fig. 1; UniFrac unweighted significance test, p ≤ 0.01)." lines 297-301.

L293 Were these ponds physically connected to each other? Or is there some other explanation for close relatedness?
The ponds sampled for this study were not connected to each other by surface runoff or streams. We concluded that the phylogenetic clustering observed is mainly related to

the ambient environmental variables that select for closely related OTUs. We have now noted the lack of connectivity (lines 477-482).

L295 Could you please state the indices that were used to estimate species richness and diversity?

We have now stated this as follows: "We then investigated the phylogenetic diversity of these communities as defined by Faith (1992) along with other diversity metrics such as the phylogenetic species richness and evenness (Helmus et al. 2007)…" (lines 193-196).

L319 How reliable is this classification? Are you sure it is Opitutus, or is it a bacterium closely related to it? How were the oxygen conditions in the ponds? Methods only state "surface water", which specific depth is this and were the oxygen conditions comparable among the ponds? Oxygen is a major determinant for microbia, and if some of the ponds have oxic and others anoxic conditions in the sampled depth, how do you justify mixing these different redox conditions in a single co-occurrence analysis?

We agree with the reviewer that oxygen is an important environmental variable driving community composition but also ecosystem function, especially in term of methane dynamics. As presented in the new SI Table 2, dissolved oxygen was variable among ponds but no significant difference was identified among the different valleys (1-way ANOVA, P>0.05) such that no valley presented on average lower level of oxygen. To address this concern, and given the low percentage of sequence similarity, we have removed any mention of the genus *Opitutus*.

L326 I was not able to interpret which OTUs in table 2 are these discontinuous and sporadic –associated OTUs?

We have revised Table 2 in order to improve its clarity. Specifically we highlighted in bold the different permafrost landscape types and thus the corresponding inhabiting OTUs.

L332 How big proportion of the community did these OTUs comprise?

We have added this information in the revised text (lines 347-353), which now reads:
    "In particular, OTUs assigned to methanotrophic bacteria representatives were prominent within the sporadic permafrost landscape type: OTUs closely related to *Methylotenera* (OTU 10) and *Methylobacter* (OTU 9) were among the five most abundant taxa at SAS sites (3.5 and 3.6 % of the total number of SAS reads respectively) and OTUs assigned to methanotrophic Verrucomicrobia *LD19* (in the class

Methylacidiphilae) was one of the most abundant OTUs at the KWK site (Fig. 2, 1.4 % of KWK reads).”

L341 Do you have any explanation why the community in BGR appeared to be randomly distributed? Any environmental factors that could explain this?

It is unlikely that environmental variables could explain this pattern. BGR valley was not identified as the valley with the highest level of environmental heterogeneity, despite the fact that SI Table 2 shows that local environmental conditions vary among ponds within BGR valley. This in turn may indicate that in this valley, neutral processes may be primarily involved in bacterial community assembly as opposed to other valleys where deterministic, environmental filtering related, processes may shape the local communities. However, we do not have enough information and would prefer not to speculate on this in the text.

L347 Sentence starting with “In general. . .”. I don't understand what the authors' message in this sentence is.

In the revised manuscript we have removed the text mentioned by the reviewer and this section (lines 366-369) now reads:
   “The OTU co-occurrence patterns as well as the relationships among both biotic and abiotic variables were investigated by network analysis. The most connected nodes (degree > 10) were related to three abiotic variables (DOC, conductivity and TP) and one biotic variable (autotrophic picoeukaryotes).”

L364 What about relationship with the actual abundance of these OTUs? I think the number of organisms within each of these OTUs is much more relevant than the number of OTUs from the same phylum. Were the most connected OTUs also the most abundant ones?

We thank the reviewer for this suggestion and have now added the new Figure 7 in the supplementary information that presents the relationship between the degree of OTU nodes from the dominant bacterial phyla and their abundance (% of reads * total bacterial abundance, cells ml^{-1}). The corresponding text in the Results (lines 399-401) is as follows: “In general, our results indicated that the most connected OTUs were also the most connected ones (R^2=0.25, P<0.001, SI Fig. 7) even though some of the most connected nodes (OTUs) had low abundance (SI Table 3).
We also added the following text in the Methods (lines 278-283): “The relationship between the connectivity of OTUs (as indicated by the degree value in the network) and their corresponding abundance was examined in generalized linear models in order to relax from normality assumptions. OTU abundance was first calculated per individual pond as the product of % of total reads and total bacterial abundance. The total

abundance of an OTU in the dataset was then obtained by summing the abundance calculated for each pond".

L381 Could the authors please clarify how the reduced network was constructed. The text mentions removal of the most connected nodes while the table legend says "network centered on the 24 most connected OTUs".

We have rephrased the caption of Table 4, which now reads: " […] whereas Hubs refers to a network where the most connected 24 OTUs from the whole network (SI Fig 5A) were removed prior to this analysis. Grey shading refers to topology characteristics of Erdős–Rényi random networks of similar size.

L432 DOC concentrations?

We have changed "DOC content" to "DOC concentration" as suggested by the reviewer.

L435 The order in which these valleys are on the permafrost gradient is very important for interpreting the results. Please include it to a table.

We have now added a new Table 2 in supplementary information that shows the permafrost landscape type for each pond. We also now refer to a recent companion study (lines 103-104) that presents the location of the different valleys along the permafrost gradient.

L474 Overdispersed?

This sentence now reads: " When controlling for the two outliers mentioned above (NAS-A and NAS-B), the northern communities (BGR, NAS) presented a greater phylogenetic distance among co-occurring species than expected by chance (lower NRIs) than communities from the thaw ponds located in valleys from sporadic permafrost (KWK, SAS)."

L515 Again, how were the O2 conditions among the ponds? Is it realistic that this OTU really is an anaerobe? Or just closely related to one?

Dissolved oxygen concentrations are now presented in SI Table 2. Given the uncertainty regarding taxonomic assignment at fine resolution as discussed in an earlier comment from the reviewer, we have removed the sections that specifically mentioned the bacterial genus *Opitutus*.

L563 "The bacterial networks similarly showed differences among valleys, corresponding at least in part to differences in DOC and conductivity." Maybe I missed something, but I got the impression that there was only 3 networks; based on real data, reduced real data and a random network. Also table 4 shows data only for these 2. Now I'm a bit confused which are these networks that show differences among valleys?

We thank the reviewer for pointing this out and have revised the text : "The bacterial networks further showed that DOC and conductivity played an important role in the structure of co-occurrence patterns of bacterial OTUs, corresponding at least in part to differences in these two environmental variables among valleys (SI Table 2)" (lines 608-610).

Figure 1. A legend illustrating the colors in panels a and b would improve the readability.

Following reviewer's suggestion, we added a legend for color code in Figure 1 and revised the figure caption, which now reads: "(A) UPGMA clustering based weighted and normalized UniFrac distances among bacterial community samples. Clustering statistics were computed using 100 jackknife replicates. (B) Principal coordinate analysis (PCoA) using UniFrac weighted distance metric. Valleys SAS and KWK are located in the sporadic (highly degraded) permafrost landscape, valleys NAS and BGR are loicated in the discontinuous (less degraded permafrost landscape, and reference rock-based lakes are located in the RBL valley. (C) Differences in the phylogenetic structure (NRI, net relatedness index) of bacterial communities among the different valleys. The solid black horizontal and vertical lines represent the mean and SD respectively. The dashed line represents the mean NRI value of NAS valley, with the 2 outliers excluded. Black dots represent individual pond and lakes".

Table 2. Legend and data in conflict regarding numbers presented for each site. Legend states top 10 per site. How ever, NAS only has 9 taxa while SAS and KWK have 14.

In response to the comment, we have revised this caption as:
 "Results of indicator species analysis. Valley refers to the valley (or combination of valleys) for which the OTU obtained the highest correlation. We indicate the value of the correlation (r.g) and its statistical significance (P) at α=0.05. Only OTUs with r.g ≥ 0.6 are presented when associated to one valley (only top 10 is presented for KWK and SAS valleys). OTUs were classified at their finest taxonomic levels based on similarity to sequences in Genbank database."

Table S1. Does number of taxa correspond to number of OTUs?

Yes, we here stated this more precisely as requested, and replaced "taxa" by OTUs.

Anonymous Referee #2

General Comments
I thoroughly enjoyed reviewing the manuscript from Comte et al. describing the diversity of the bacterioplankton and the patterns of co-occurrence among bacteria, phytoplakton, autotrophic phytoplankton, and zooplankton in thermokarst ponds across five valleys along a North-South permafrost degradation gradient in Nunavik. This study set out to determine whether there were habitat preferences shown in the diversity and co-occurrence patterns of bacteria along the permafrost gradient and, if so, whether these preferences related to differences in phylogenetic structure. This was accomplished by using 16S rRNA gene pyrosequencing to describe the bacterial community, and the collection of physico-chemical and biotic measurements, along with the characterization of the phytoplankton and eukaryotic plankton community in the lakes. The authors found differences in bacterial communtity composition across valleys strongly related to landscape type, but also showed a high degree of taxonomic relatedness within and across valleys, due to a few dominant OTUs. They also showed strong local effects (pond-by-pond) on the beta-diversity of the bacterial communities and were able to identify indicator OTUs for each of the landscape types and dissolved organic carbon, phytoplankton biomass, and salinity as the main explanatory variables for the changes in diversity. The authors argue that these lines of evidence suggest habitat preference on the local scale, and in the less abundant bacteria, but with a higher similarity among the highly abundant bacterial OTUs across the landscape types. They suggest this is driven by the high variability within valleys and the DOC and salinity concentrations found in the lakes. The co-occurrence networks did find distinctions between the valleys in co-occurrence patterns, and also pointed to DOC as having a significant, if complex relationship to the bacterial co-occurrence patterns. The co-occurrence networks also suggested potentially important inter-relationships among the bacteria and identified highly connected, potential "keystone" OTUs in the network. Overall, I think this paper was clearly written, the study was carefully done, and the data was thoroughly analyzed. I think that this study provides an interesting look into the patterns of microbial community assembly in spatially separated thermokarst lakes and provides insight into the assembly rules that generate these communities. I have a few suggestions, but I do not think that they involve any major changes in the manuscript, and I have not identified any fatal flaws. I think that these results are interesting: that the main differences in the microbial ecosystem are taking place among the less abundant OTUs, including inferred interactions and keystone OTUs. I commend the authors for a study well done.

We thank the reviewer for these positive comments.

My main concern is that MIC lacks sign, or direction, so it does not provide information on whether there is a positive vs. negative correlation (or co-occurrence vs. non-cooccurrence). In the Reshef et al. 2011 paper where MIC is described, some of their strongest relationships identified in their mice microbiome datasets were what the authors referred to as "non-coexistence". Using MIC allows for the identification of many of these relationships, but without looking at the relationships it's tough to interpret what they mean. The authors do some interpreting, particularly with

Chitinophagaceae, but it would be strengthened if they showed examples of these relationships to bolster their point. I believe the authors have found potential relationships that point to ecological function, I just want to know what kind of relationships they've found. This would also allow the authors to suggest further hypotheses regarding function that could be tested in these or similar systems.

We thank the reviewer for this comment, and have now incorporated examples of link between OTUs detected in our network analyses with high MIC values (strong relationship) but which differ in their linearity. The non-linearity metric allows distinguishing between co-occurrences (low non-linearity value) and non-co-occurrences (high nonlinearity value). In particular, we illustrate linkages between OTUs assigned to Chitinophagaceae and Actinobacteria. In addition we show examples of co-occurrence and non-co-occurrence between OTUs (Fig.5).

Specific Comments
I would find a map or at least a schematic of the sites a helpful addition either in the main text or in the supplemental materials.

This work builds on an earlier study that is now published and that contains a map of all sites. We now cite this in the revised manuscript (lines 103-104).

In some instances, the authors seem to be arguing for the taxonomic ID of OTUs to suggest ecological function, when they have evidence in their study which would allow them to infer function. I think this can be resolved through a change in emphasis in the discussion, rather than any major analytical change. The identification of potential keystone species is interesting in these thaw-ponds and lakes, but there is a lack of detail about likely function that I believe the authors may be able to infer from their networks and data.

Reviewer1 had a similar comment that we should be very careful in linking 16S data to specific functions, and we agree with the reviewers. In the revised version of the manuscript and following the suggestion of reviewer2, we focused on the OTUs that were identified as hubs ("keystone species") and investigated whether they were closely related to known bacterial taxa that belong to certain functional groups. Although most of the bacterial hubs in this study were related to typical freshwater generalists (e.g. *Rhodobacter, Limnohabitans*), there were also OTUs related to specialists such as nitrogen fixing bacteria (*Beijerinckia*) and chitin degrading bacteria (*Sediminibacterium*) as indicated in SI Table 3.

I also suggest that the authors emphasize that the "keystone" OTUs identified as hubs were not merely the abundant OTUS, but were rare, potentially important actors of the particular ecosystem. This might be worth pointing out in the abstract.

As suggested by the reviewer, we have noted this point in the abstract. This comment also relates to the query from reviewer1 who asked to relate the connectivity of nodes (OTUs) in the network and the abundance of the corresponding OTUs. In response to this, we found a significant positive relationship between the number of edges of each node (bacterial OTU) and its abundance within the dominant bacterial phyla (SI Fig 7). This in turn suggests that the most connected OTUs tend also to be the most abundant ones. However, there were some exceptions as some the hubs were in fact low abundant as indicated in (SI Table 3). We have added more information on this point in the revised text.

In addition it may be worthwhile to look for any overlap in indicator OTUs for the different habitats and highly connected nodes. I think this may provide deeper insight into these thermokarst microbial ecosystems. It's possible the data is not there for this, but the authors may be able to suggest ways to resolve these types of connections.

We thank the reviewer for this suggestion and have now have added the node degree information on Fig. 2. This shows that OTUs with some degree of habitat preference among the most abundant OTUs were also among the most connected ones. In particular, the figure shows that OTUs that were mostly detected in southern most degraded permafrost valleys, were also highly connected. In contrast, OTUs that were mostly detected in northern less degraded valleys, presented a low level of connectivity. We have discussed the implications of these findings in the revised text. In addition, in panel B we constructed a second heatmap that is based on the 30 most connected OTUs (in contrast to panel that based on the 30 most abundant OTUs). The figure shows that the most connected OTUs also show some habitat preference and that 14 out of the 30 most abundant OTUs were also among the most connected ones. This is also noted in the revised text at lines 571-579.

Technical Corrections
Pg 11 L17 The sentence "The discrepancy between dw4000 and duw4000 patterns. . .of a few but highly abundant OTUs within different valleys" Is a bit confusing. I think replacing it with ". . . of a small number of highly abundant OTUs within different valleys." is clearer and doesn't change the meaning. But in any case it should be re-worded.

We have rephrased this sentence as suggested.

Pg 19 L 23 "autrophic eukaryotes" should be autotrophic.

Thank you for spotting this typo, now corrected.

Table 3. I cannot tell the difference between the text in bold and the normal text. I see the p-values, but not the bold distinction

We apologize for this inconvenience. The text in bold appeared on the manuscript we submitted to BG but did not appear in the formatted pdf version from BGD. We will check this carefully at proof stage.

Table 4. I can't see the grey shading mentioned in the figure legend.

The grey shading was obvious in the material we submitted to BG but was not in the formatted pdf version from BGD. We will check this also at proof.

Figures 3 and 4 . . . Node size is not described in the legend. In the Supplemental section it's related to degree, but it should be stated here as well.

We have now added this information to the figure captions.