**Response to reviewers related to Linz et al, mSphere 00169-17**

Reviewer #1 (Comments for the Author):  
  
The authors characterize the bacterial communities in the epiliminion and hypoliminion of eight bog lakes with different mixing regimes and appear to be addressing the primary hypothesis that mixing regime is associated with bacterial diversity. While the results demonstrate site- and layer-specific differences in bacterial communities, the hypothesis is not well addressed throughout the results. Furthermore, the interpretation of the results is difficult as the text is currently laid out -- in many sections it seems the authors are describing software outputs without sufficient contextualation or interpretation.

*We have added better contextualization throughout the results section, with care taken to begin each section with the motivation for each analysis and conclude with what was learned from that analysis. The primary hypothesis (mixing associated with diversity) has been better linked to each analysis where appropriate; however, analyses addressing the variability and the core community of these systems are not directly related to the primary hypothesis. We did not initially expect the observed levels of temporal variability and felt it was important to include in conjunction with our research on diversity and mixing. Rather than retroactively modifying our hypothesis based on the data, we instead report surprising results that do not directly address the impact of mixing.*

It is also particularly troubling that most of the results are based on PCoA analyses, although the description of beta diversity calculations seems inaccurate.

*The reviewer makes an excellent point that PCoA analyses and other ordination methods are primarily exploratory/visualization exercises. To strengthen our conclusions, we have added supplemental heatmaps showing average beta diversity between sites (Figure S5, g-h). However, we would argue to keep the ordinations in the main text as we feel they provide a better visualization of the same results, improving reader understanding. We have also clarified the description of beta diversity calculations in the methods (L487).*

As the authors themselves mention in the discussion, this work could be furthered considerably by more careful and deliberate analyses of the data (AND METADATA!) to specifically address their hypotheses.

*When we stated in the discussion that* “We hope that this dataset and its future expansion will be used as a resource for researchers investigating their own questions about how bacterial communities behave on long time scales” *(L425),* *we meant that our dataset could be used in conjunction with other data or that highly specific questions about freshwater taxa could be asked using our dataset, not that others should perform analyses that we had overlooked. We were careful and deliberate in addressing a set of research questions that we found to be most compelling in this first analysis of this dataset. The limitations of this dataset (irregular sampling intervals, irregular sampling of environmental data, different numbers of samples in each site and layer, and large number of confounding variables in the environmental data) present many statistical issues (e.g. as considered by Reviewer 1 in Fig 3d). As Reviewer 2 mentioned being unaware of these limitations until the discussion, we have added sentences in the first results section with this information (L134) and added a figure representing the time, location, and paired data for each sample included in this dataset (Figure S1).*

*We feel that including all the environmental data would make this manuscript long and unfocused, but we have added such data where appropriate to describe general features of the lakes. This includes a visualization of water column temperatures in conjunction with richness over time in Fig S3 and alternative colorations of Fig 2 representing environmental variables in Fig S5.*

While I believe this dataset to be of value in understanding long-term community dynamics in these bog lakes, the current presentation of data does not offer novel insights beyond what is currently known in the literature. Also of note, the authors address the need to sample over a long time scale to adequately understand community dynamics and describe this as a finding of their study, however, this seems obvious and has been well demonstrated in other studies (e.g. work on the English Channel).

*To address this concern, we have added language in the introduction emphasizing the need for long-term sampling specifically in freshwater (L95). Many freshwater studies draw conclusions about bacterial community dynamics based on one or two years of data; therefore, we feel that longer time scale studies are necessary to provide context for these findings. While there are several multi-year time series in marine systems that already emphasize the need for long-term sampling, confirming this finding in an aquatic ecosystem confined within clear boundaries is still a useful result. Additionally, we saw that community dynamics did not repeat annually during summer stratification, which is different than what was observed in the English Channel, other freshwater lakes, and other similar studies (summarized in the discussion, L357).*

Specific comments:  
Abstract  
In general, the abstract describes general trends observed but very little actual information is presented. The abstract should be revised to state more specific findings. For example, which taxa were ubiquitous to freshwater?

*Specific taxa found to be core to all lakes sampled were added to the abstract (L45), as well as examples of taxa that were indicators for each lake studied (L42).*

line 36: "Multiple sites..." this sentence adds little information and should be revised or removed.

*This sentence has been removed.*

Introduction  
line 104: Change "contains" to "contain".

*This change has been made (now L113).*

line 108: This sentence is awkward and should be revised.

“Our primary goal for this dataset was to census the bog lake community and determine which taxa are core to all bog lakes, to each thermal layer, and to each mixing regime” *has been revised to* “Our primary goals for this dataset were to census members of the bog lake bacterial community and to identify taxa that are members of core communities in bog lake ecosystems” *(now L118).*

Results  
line 129-131: It is awkward to discuss Figure 2 prior to Figure 1. Perhaps this statement can be moved to methods to avoid distraction from results?

*We have moved this sentence to the “Statistics” section in Methods.*

line 136: Please provide the statistical test used to evaluate differences as well as the p-values.

*Information about the statistical test used has been added to this line (now L154). P-values are now provided in Table S1.*

line 164: The axes on an ordination plot are dimensionless, thus the disperson the authors are addressing is not caparable between ordination plots. Furthermore, the numbers of samples involved as well as the numbers of years included will affect the ordination. Thus the results presented here are largely artifacts of ordination and not discrete measures of variability. The authors should provide some justification for this methodology or remove it.

*The reviewer is correct that comparing metrics between ordination plots is invalid, but this is not the analysis that was performed. Dispersion was compared between lakes within the same ordination plot, specifically, the one presented in the previous Figure S4 (replaced to make room for new supplemental figures based on reviews) . However, because it is true that artifacts can impact the results of ordination, we have chosen to compare pairwise weighted UniFrac distance within sites in Figure 3d instead. To address concerns about different numbers of samples taken from each site, we also ran this analysis on only data from 2007, when all sites were sampled approximately equally. This analysis shows the same results as those obtained from the full dataset (L210).*

line 184: What is being presented in parentheses? If these are taxonomic classifications, more detail should be provided (e.g. genus).  
line 195: Similar to the previous comment, these tribe affiliations are relatively meaningless. The authors should provide more easily interpretable taxonomic information.

*We have clarified in the text that the items in parentheses are taxonomic classifications from the provisional but hierarchical freshwater-specific 16S rRNA gene taxonomy database, published in Newton, et al “A guide to the natural history of freshwater bacteria.” 2011 (now L225). The lineage-clade-tribe structure is approximately equivalent to 90%-95%-97% sequence similarity in the full-length 16S region (though also based on careful phylogenetic reconstructions) and provides finer resolution than a genus-level classification. This taxonomy is definitive and well-accepted in the freshwater field; it has been cited 349 times since its publication in March, 2011.*

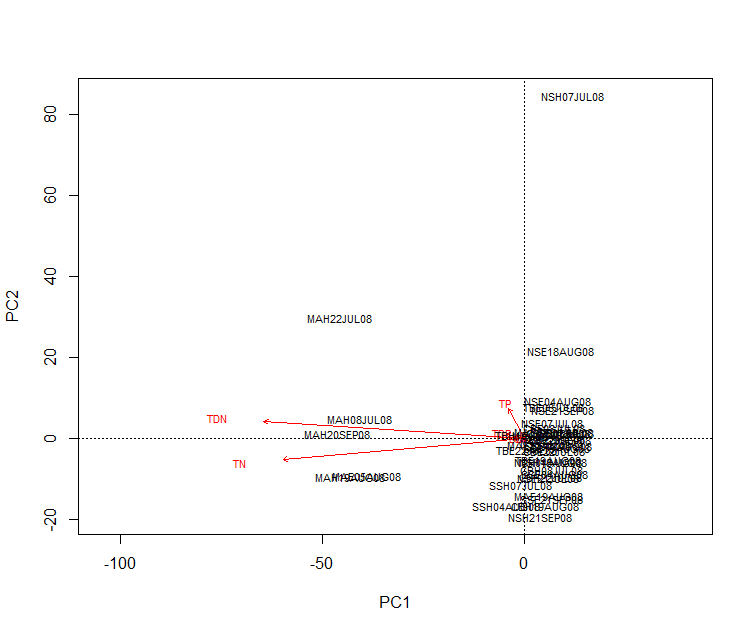
Discussion  
line 259: A word is missing here.

*This typo has been fixed (L317).*

line 281: It is not clear how the results address this hypothesis.

*A statement clarifying that mixing is the disturbance being tested has been added to the discussion (L305).*

line 283: Why were these metadata not included in any analyses? It is a fairly straightforward analysis to determine to what extent mixing vs. physicochemical parameters are influencing the community, for example, by variance partitioning using partial redundancy analysis.

*We thank the reviewer for this helpful suggestion. We performed partial redundancy analysis and determined that nitrogen concentrations explain differences between samples, particularly in the Mary Lake hypolimnion vs all other sites. However, because only a limited number of samples (75) have these measurements, and they are from only late summer/early fall of 2008, we have chosen not to add these data to the main text because we cannot assume that it is representative of the entire dataset. We include these environmental data in Table 1 to give readers a broad sense of the chemical limnology of these lakes, which may be very different from other lakes that readers are familiar with.*

line 310: Please change "names" to "taxonomic classifications" or something more meaningful.

*This change has been made (L382).*

Figures  
The authors have a tendency to over-interpret figures in the captions, making the figures themselves difficult to understand. Please revise captions to describe figures and reserve interpretation for the text of the results section.

*While we have made reductions in figure legend length, particularly in areas of interpretation, we prefer that figure legends contain enough information to stand on their own without reference to the main text (L710 – 763).*

Figure 1: This figure is extremely confusing as presented, especially in regard to statistical significance. Please revise this. Perhaps use letters to denote significant differences.

*We have switched to letters instead of colors to indicate significant differences.*

Figure 2: Please provide r-squared values for the axes shown.

*We are unfamiliar with r-squared values for principle coordinates analysis axes, but did report the percent variance explained by each axis on the plot, a metric frequently used in the literature to assess the fit of an ordination. However, we realize this was likely unclear. We have better labeled the percent variance on the plot and included a note in the legend (L723). R-squared values for the PERMANOVA testing significant differences by lake and mixing regime are now included in Table S2.*

Supplemental material  
Table S1 is redundant with figure 1 and is not needed.

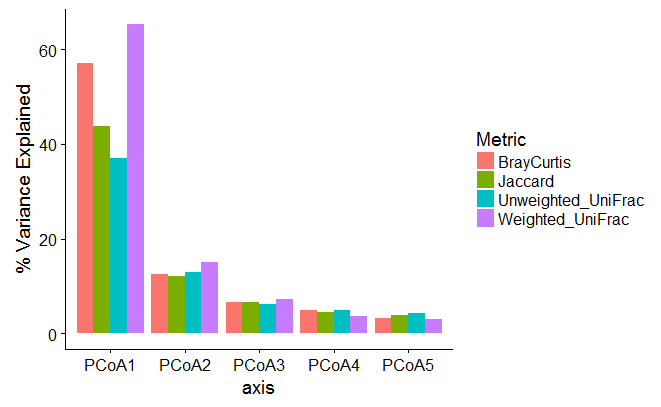
*Table S1 has been removed to provide space for new supplemental figures and tables.*

Methods  
line 547: Please clarify what is meant by "biological replicates" here. It is not clear how sampling and replication was handled as written.

*Clarification on how biological replicates were collected has been added to “Sample Collection” in the Methods section (L435).*

line 402: Please clarify what parameters were used for beta diversity analyses and ordination. Bray-Curtis and Jaccard are alternative measures of similarity to UniFrac; how were these matrices used to compare Unifrac distances?

*The description of beta diversity metrics and their comparison has been clarified in the text (L487). The results of that comparison are presented here.*

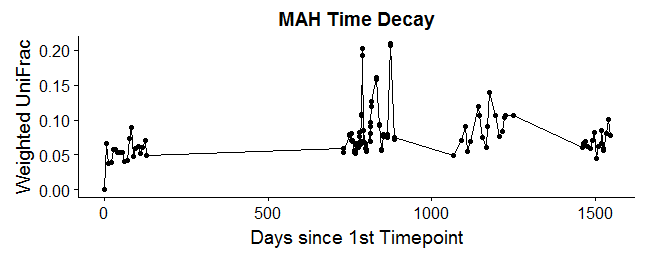


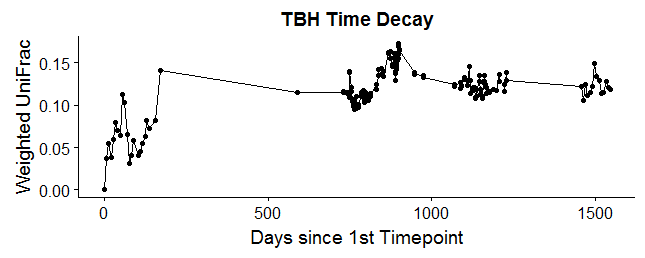
Raw data must be made available through an independent sequence repository such as the Sequence Read Archive at NCBI.

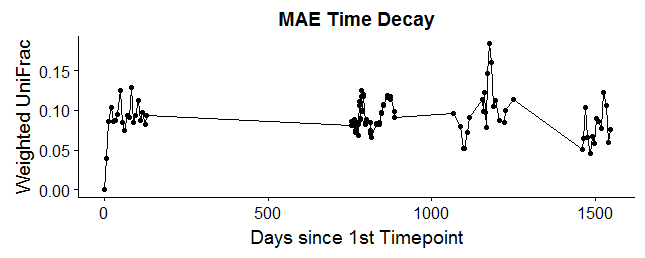
*We thank the reviewer for pointing out that we did not include information on the location of raw data. Raw data is available through QIITA and the independent sequence repository at the European Bioinformatics Institute. We have added this information to the “Data Availability” section (L514).*

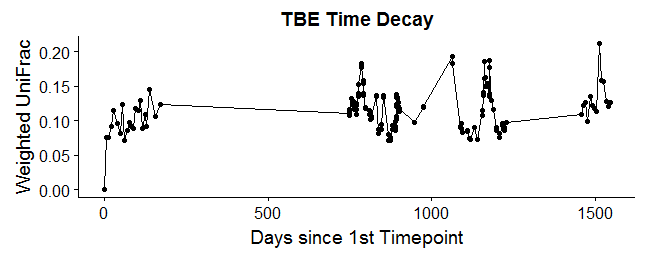
Reviewer #2 (Comments for the Author):  
  
Linz and co-authors describe an interesting community dynamics time series of 8 bog lakes over multiple years. They sampled the epilimnion and hypolimnion and performed 16S rRNA gene sequencing. They tried to address several questions regarding the presence of a core community, the reproducibility of community dynamics over years, and of a relationship between mixing regime and richness as a test of the intermediate disturbance hypothesis.  
  
Overall, this study is a valuable contribution as multi-year time series data are sparse for freshwater systems. I do feel some more in depth analyses would significantly strengthen this study however.  
(a) betadiversity compared to starting point over time. I believe Jack Gilbert presented such analyses in his ocean time series work, and I guess the meta-analysis did as well. Would still be worthwhile to repeat here in depth. This is the main test to evaluate whether there is some temporal reproducibility in community composition

*We have previously performed beta diversity analyses investigating changes in community composition vs the time zero sample and found no support for repeating annual trends. Unfortunately, there is not enough space in the supplemental documents in for the results of these beta diversity time decay analyses, but we have included a statement reporting the absence of a trend in the text (L191). A few example plots are included below:*









(b) evaluation of whether the pattern differences are due to differences in env parameters changes between years, (Temp, PAR, ...), or are stochastic, or are due to founder effects of differences early on in the year that set the communities on a different track every year. It is unclear to me whether environmental data was measured to accompany the microbial data... would be a bummer if not. From the methods it seems there should be data available for at least some lakes and temp for all of them. An analysis of whether differences between years can be linked to differences in measured variables would be valuable and should be done.

*We agree with the reviewer that this would be an interesting analysis to conduct; indeed, there are multiple variables that vary by year. However, with only four years of sampling, we still cannot link community types to environmental parameters. Given the extent to which many different variables change across years, we estimate needing 10 years of sample data to adequately link environmental data to community composition. (Kara, et al. "A decade of seasonal dynamics and co-occurrences within freshwater bacterioplankton communities from eutrophic Lake Mendota, WI, USA." The ISME Journal 7.3 (2013): 680-684). As an example, here are mean values from Trout Bog of potential drivers of interannual variability in the bacterial community:*

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | Ice-off date | Daily High | Daily Low | Rainfall | Epilimnion Temperature | Hypolimnion Temperature |
| 2005 | 14-Apr | 66.71 | 43.5 | 0.11 | 16.78 | 5.32 |
| 2007 | 18-Apr | 54.43 | 31.67 | 0.08 | 14.49 | 5.11 |
| 2008 | 30-Apr | 60.92 | 37.81 | 0.09 | 16.55 | 5.86 |
| 2009 | 20-Apr | 60.89 | 38.55 | 0.08 | 15.85 | 5.33 |

*We also considered founder effects, but without winter/early spring samples for each lake and year, this cannot be tested. We spent a significant amount of time attempting to design an analysis to measure stochastic effects (such as using nearest taxon mean distance as a metric for the influence of stochasticity vs environmental drivers in shaping a community) but ultimately concluded that proving the absence of any environmental driver would be needed to prove that the annual variability is the result of stochasticity, a task that is not feasible at this time.*

(c) The authors claim to test the intermediate disturbance hypothesis, but clearly the richness is correlated o differences in depth. They should at least be able to test whether what they are seeing is a pattern as. function of a simple species-area relationship. They can test depth, area, volume, for the whole lake or by lake layer.

*We have previously tested the relationship between volume and richness and found that they are positively correlated. Lake volume has been added to Table 1. However, because our original study design specifically addressed the environmental gradient of mixing frequency, and many limnological variables co-vary with depth/volume, we cannot conclude that a positive taxa-area relationship exists in bog lakes. More lakes with variable volumes but similar depths would be needed to test for the presence of this relationship. We have revised our discussion of alpha diversity results to include this perspective (L345).*

Other comments:  
L. 38: likely, isn't there a lot of metadata available?

*Yes, but variables in the environmental data co-vary, making identification of driving factors difficult. See Reviewer 1’s suggestion of pRDA to dis-entangle.*

L. 39: how high compared to seasonal or inter-lake diff?

*Interannual variability is greater than seasonal variability, but less than inter-lake differences. To avoid subjective terms, we have changed “high” to “unexpected” in the abstract (L44).*

L. 40: "core" in all lakes or in each lake and layer?

*This sentence has been clarified per Reviewer 1’s suggestion (L45).*

L. 49: related to the main point b, do the dynamics really operate on multiple years or is each year diff because of (a) stochaSticity, (b) differences in env factors, etc. I do not think you can make the statement you do based on the current analyses

*This line has been revised to clarify that we do not describe multi-year trends; rather, we observe differences in community composition each year (L55).*

L. 74: but was this large or small?

*We are confused by this comment because the size of this effect is relative to the frame of comparison. Changes in beta diversity across space vs time probably depend on the physics of the lake, amount of horizontal mixing, temperature, and other lake-specific factors. This result was intended as an order of magnitude measurement only (now L78).*

L. 81: "range of potential change" --> time scales of change?

*“range of potential change” has been reworded to “range of potential time scales for change” (L87).*

L. 109: series of goals, makes the paper a little less focused - some of the goals listed here are not touched upon in the abstract, such as intermed disturb hyp

*We have improved consistency in how we describe our goals throughout both the abstract and the main text.*

L. 137: perhaps list from most to least or least to most mixing events? Some people may not be familiar with limnological jargon terms

*We have re-ordered as the reviewer suggests (L155).*

L. 157: permanova values for yearly effects?

*A PERMANOVA table for yearly effects has been added in Table S2.*

Results: I would really urge you to make a comparison of how betadiversity changes (you can represent both t vs t-1 over the years as well as t1, t2, t3, ..... vs t0 for each lake)

*Beta diversity time decay plots (t1, t2, t3 … vs t0) do not support repeating annual trends. We have also performed the beta diversity analysis comparing t vs t-1 and found that changes in beta diversity between consecutive samples (turnover) are highest during mixing events. However, as this is not a main point of our manuscript, we have chosen not to include these plots in the supplemental document for the sake of focus and concise text.*

L. 163: betadispersivity: what exactly is the point you are trying to make here?

*Per Reviewer 1’s comments about clarifying the motivation for each analysis, we have stated in the text that our point was to compare variability between sites (L215).*

L. 232: important result. Can you put this in some more context in the discussion, e.g. compared to oceans? Why similar/different

*We have added a paragraph to the discussion comparing our results relating to consistent lifestyles of freshwater taxa to findings in marine systems (L383).*

L. 257: "each year harbored": there is no yearly community, only assemblages on specific days at specific times.

*“*Each year harbored” *has been reworded to* “The community composition observed in each year of sampling” *(L313).*

L. 259: incomplete sentence

*This typo has been fixed (L317).*

L. 260: Again, seems the analysis is a bit lacking on how temporal and spatial scales affected community divergence

*As this discussion section reviews literature previously published on our study sites, we are not sure what the reviewer would like added to address community divergence.*

L. 272: is it mixing regime or differences in bathygraphy and other env factors though? These may be correlated to each other, but whether the mixing regime is causal or not is not supported by the current analyses

*The statement “*We also supported previous research on the characteristics of bacterial communities in the epilimnion and hypolimnion, and the impacts of lake mixing on these communities” *has been revised to* “We also supported previous research on the characteristics of bacterial communities in the epilimnion and hypolimnion and the association of lake mixing frequency with community composition” *to clarify that we are not concluding a causal relationship (L332).*

L. 305: you need to set up these limitations a bit better from the beginning. I was under the impression these lakes were sampled throughout the year.

*We have added a supplemental figure showing when each site was sampled and what paired environmental data is available (Figure S1). We now reference this figure and table early in the Results section (L136).*

L. 326: here you seem to indicate weekly sampling across 4 years, rather than mostly in summer?

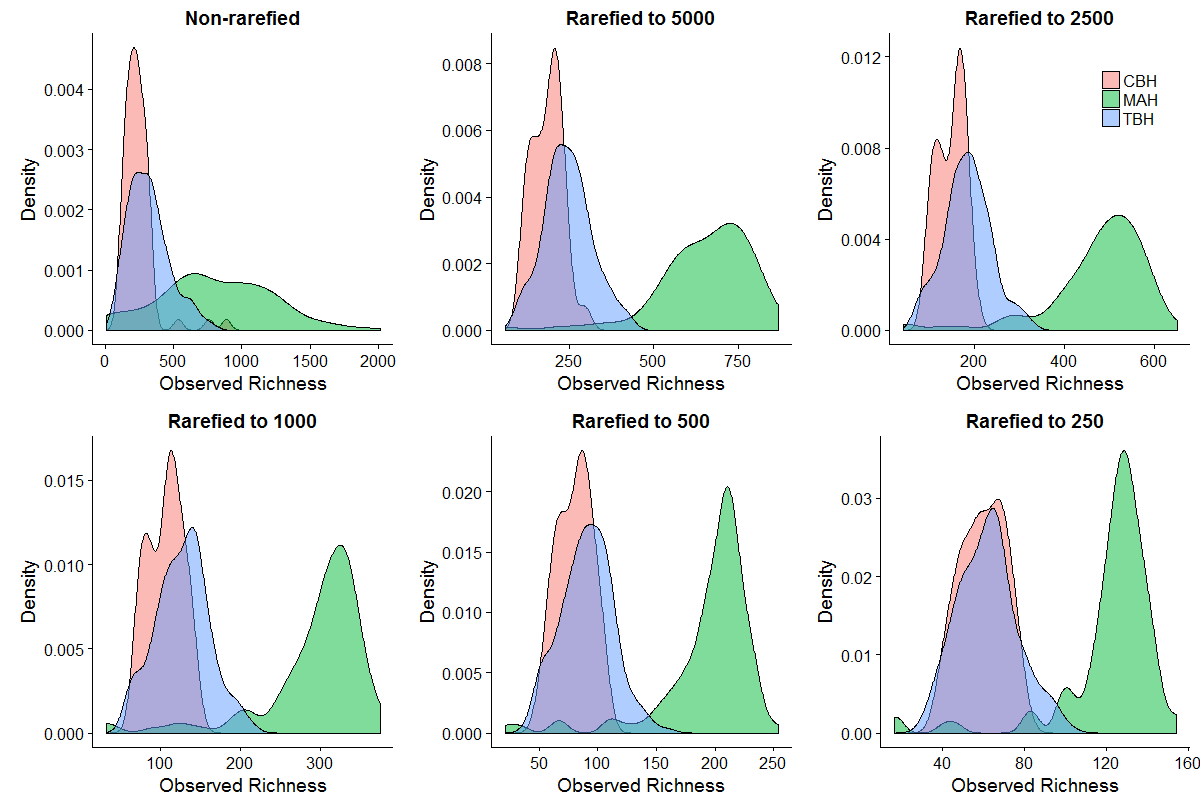
*We have re-worded this sentence to refer to summer trends specifically (L409).*

L. 373: any reason for this specific number?

*Removing reads that appeared less than 25 times in the raw sequencing data was recommended by the developers of the deblur program and is standard for their analyses of other Earth Microbiome Project datasets (L456).*

L. 379: that's a pretty low number for the richness analysis - Making me feel you should only look at inv simpson (see Haegeman, ISMEJ 2013) -- same on L. 381-383: "simulation... sufficient quality"--> and how do you assess this? See Haegeman, who would argue 2,500 is way too little to be able to say anything meaningful about richness.

*We appreciate the reviewer’s suggestion of this paper. The inverse Simpson index, the Shannon index, and the Chao1 metric were all previously tested as potential indicators as richness, and all showed that richness varied with depth/mixing regime. However, observed richness and evenness (measured by Pielou’s evenness index) were strongly correlated in our dataset. Because the Simpson and Shannon indices include information about evenness, we chose not to use these indices to unlink those two properties. The Chao1 metric uses the number of OTUs appearing with only one or two reads in the entire dataset to estimate “true” richness and does not include evenness information; however, we were concerned that this metric was highly sensitive to read abundance thresholds imposed early in the data processing steps. Therefore, we chose observed richness as the least biased way to compare richness between sites.*

*When choosing the rarefaction cutoff, we were faced with a choice between a higher number of reads per sample but fewer samples (as samples with less reads than the rarefaction cutoff are removed from the dataset), or a lower number of reads per sample and more samples. 2500 was chosen based on the results of simulations to maximize the number of samples retained while losing a minimum of observed OTUs to subsampling (L464). 2500 is used in other similar studies and we believe that, while on the low end, is still a reasonable number. Additionally, we investigated using non-rarefied proportions of read abundances instead of rarefaction and obtained similar results. A comparison of richness between three sites (CBH, TBH, and MAH) at different rarefaction levels is included here to demonstrate that our main finding, that richness varies by site in rank order, holds regardless of the chosen rarefaction cutoff. The point presented in Haegeman, et al that less diverse communities appear more diverse at very low rarefaction cutoffs appears to be a concern only below a rarefaction cutoff of 500.*

L. 592: same day and time of day?

*A statement explaining that nutrient samples were collected concurrently with bacterial samples has been added to the table legend (L699).*

Table 1: How meaningful is a single value considering how much these numbers fluctuate across the season (and to some extent years): see Hanson et al, Ecol Monogr 2006

*While these numbers do fluctuate, our intent is to merely demonstrate the differences in nutrient concentrations between bog lakes and other systems to readers familiar with nutrient levels in freshwater. We have added standard deviations to Table 1 to make readers aware of the spread of these measurements (L700). Coincidentally, our nutrients were measured by the authors of Hanson et al, Ecol Monogr 2006 :)*

Figure 1:  
Can you color them by lake type?

*To aid reader comprehension, each lake is represented by the same color in every figure. We do not wish to confuse readers by using a different color in one figure only.*

-so is it depth or mixing that leads to correlation with diversity - depth would be a simple species-area type relationship? Did you evaluate this (volume, depth, area)? So may have nothing to do with disturbance regime? Perhaps linear models with depth, area, volume (of epi and hypo separately) and richness would be interesting.

*Please see above comments relating to Reviewer 2’s point c.*

-L.607: based on what stats test?

*The significance test used, PERMANOVA, is listed on L724.*

-again, see Haegeman regarding use of richness - your observed richness may be meaningless in relationship to your community richness

*Please refer to above comments regarding richness (L464).*

-just using letters to group samples not significantly diff from each other would be easier I think

*We have changed how we report significance in Figure 1 to letters.*

Figure 2: I would rather see a PERMANOVA table than this figure to be honest... All it shows is that the lakes are diff. Doesn't have coloration based on time of the year. or temp, or any other factor that may separate points.

*In addition to a PERMANOVA table in Table S2, we have added versions of Figure 2 colored by lake type, date, and mean water column temperature in Figure S5.*

Figure 3: again, color by mixing regime would help. Also, I wonder if this is the most informative vs suggested betadiv over time suggested in main comments

*Because beta diversity between any given sample and the initial sample (t0) levels off in less than one year, we believe Figure 3 to be more informative regarding interannual variation.*

*As with Figure 1, we wish to consistently use the same color for each lake throughout the manuscript to add reader comprehension.*

Figure 5: these patterns are very interesting

*Thanks! We think so too!*