ASLO
Association for the Sciences of Limnology and Oceanography

K. David Hambright, PhD · Editor-In-Chief, *Limnology & Oceanography*
University of Oklahoma · Department of Biology · 730 Van Vleet Oval, 304 Sutton Hall · Norman, OK 73019 · 405-325-7435
Rachel Yehl · Managing Editor, *Limnology & Oceanography* · lo-edoffice@wiley.com · 708-955-5501

RE: LO-19-0096-Time-series metatranscriptomes reveal conserved patterns and ecological interactions between phototrophic and heterotrophic microbes in diverse freshwater systems

Dear Dr. Linz:

Thank you for your submission to *Limnology & Oceanography*. Based on the reviewer comments, appended below, we ask that you revise your manuscript and resubmit it to us for further consideration. We consider this a request for MAJOR REVISION.

Both reviewers agree that this study presents a very interesting dataset, but they also raise substantive criticisms on the present form of the manuscript. There are methodological details that can be improved and clarified and, most importantly, reviewer 2 pointed to the fact that this promising dataset provides little novel insight on gene expression trends. We also agree that the time series data could be better presented and exploited. So please consider these comments and try to redo some of the analyses to focus on the novel aspects of this interesting dataset.

We ask that you submit your revision for inclusion in the special issue by June 15, 2019.

Note that it is the policy of *Limnology & Oceanography* to encourage only one resubmission in response to the comments of the reviewers and editors. Generally, this resubmission will either be accepted or rejected. You should therefore take the revision process seriously and endeavor to learn from the comments, even though you may find them frustrating. Keep in mind that the reviewers and editors are trying to help you best communicate your information to the wider scientific community.

When you submit your revision, please submit a “clean” copy with all edits accepted. You should also submit a detailed statement on how you have addressed each of the comments of the reviewers and editors, including your reasons for not taking their guidance in cases where you disagree. If you find it helpful, please feel free to also upload as a third document the “edit-mode” version of your revision, in which the edits have not yet been accepted.

To submit your revision, log into https://mc.manuscriptcentral.com/lo and enter your Author Center. You will find your manuscript title listed under "Manuscripts with Decisions." Under "Actions," click on "Create a Revision." Your manuscript number has been appended to denote a revision. Please DO NOT upload your revised manuscript as a new submission.

IMPORTANT: Your original files are available to you when you upload your revised manuscript. Please delete any old files that are no longer part of your revision before completing the submission.

Thank you for your submission to *Limnology & Oceanography*, and we look forward to receiving your resubmission.

Sincerely,

Ramon Massana Molera
Special Issue Guest Editor
Limnology & Oceanography

Marguerite Xenopoulos
Deputy Editor-in-Chief
Limnology & Oceanography

REVIEWER COMMENTS TO AUTHORS:

Reviewer: 1

Comments to the Author

This manuscript by Linz and colleagues reports diel patterns in gene expression (via metatranscriptomics) linked to taxon and functional gene abundance and biogeochemistry in three freshwater lakes over two days. Lakes were sampled 12 times at 4-hour intervals over a two-day period. The trends are interesting, though not surprising (generally, more photosynthesis during the day). A third sampling day would have helped to bolster these analyses, and replicates by lake type or three of the same type of lake would have helped too. As it is, this is n=1 for each type of lake, making lake-to-lake comparisons difficult. Still, this is a relatively thoroughly sampled dataset, considering the labor and expense involved in collecting and analyzing these types of data.

There is not a very clear visual presentation of the day vs. night trends within and among lakes. The closest representation of this is in Figure 3, which shows some of this information for one lake. Given that these diel gene expression trends seem to be the key point that the authors wish to communicate, I recommend making these trends more visually clear, e.g., with an additional figure or additional figure panels added to Figure 3 (more details below).

I commend the authors for placing detailed laboratory protocols in the supplement.

Specific comments:

Title: Given that the "ecological interactions" are speculative parts of the discussion, as opposed to conclusions from clearly presented results (see below), please remove this part of the title.

Ln 193-194: Hopefully, this was also the seed sequence for the cluster. As in, hopefully the contigs file was sorted in descending order of contig length prior to clustering with CD-HIT. Either way, please indicate here whether or not this was done.

Ln 197-201: I do not understand this description of how taxonomy was assigned. Was the best hit for each coding region considered, and then some "most frequent" taxonomy identified for the whole contig? In the case of relatively complete genome bins, a more standard approach would have been construction of a phylogenetic tree from single-copy genes. Was this not possible due to relatively incomplete bins in most cases? Please clarify this section.

Ln 218-219: Please provide exact numbers for these "many samples from day two in the TB time series [that] failed to meet quality control standards," and do you have any insights as to why these particular samples might have failed more often than others?

Ln 231: Maybe this acronym was already spelled out, but I couldn't find it in a quick scan back through the paper. What is PAR?

Ln 265-271: Please place these clade names in larger taxonomic context.

Ln 305-306: This is not particularly obvious in the data in Tables 2-4. It could be because of the different categories of sugar transporters (i.e., some categories show differences while others do not, and the categories that do differ by lake). A more useful description in the text would be how many and which categories of sugar transporters exhibited significant diurnal cycles for each lake. Some of this information appears in subsequent sentences in this section, but I recommend reworking this section to make it clear that sugar transporters were not always significantly different day vs. night, and in fact, many were not.

Ln 345-346: There is no presentation of data on potential metabolic handoffs within the Results section. Is the implication that the authors plan to speculate on this in the Discussion section? If so, please make that more clear, otherwise delete this.

Ln 361 and 376: Change "genes" to "some genes" (see comments above related to Ln 305-306)

Ln 364-375: As mentioned above, you have n=1 for each lake type with no within-type replicates. At a minimum, that should be mentioned as an explicit caveat here, but I recommend deleting any interpretations related to comparisons across lake trophic statuses.

Ln 407-408: If you are going to evoke the "Black Queen Hypothesis," please explain what it is.

Figure 1: Without color, this is not much more useful than a table. The reader needs to consult the list beside each figure to understand the order of points, and there is no easy way to visually compare between lakes. Perhaps journal color charges are part of the consideration here, but if not, please color the points in each graph (e.g., one color each for photosynthesis-related, sugar metabolism-related, and other, or whatever groupings are most meaningful). I may have missed color figures elsewhere in the reviewer packet, but they were not in the bundle with main text, figures, and supplement. It would also be useful to number the non-autotrophic genes in panels D-F by their actual rank numbers, instead of 1-10. Given that hypotheticals/unknown function can exist in both lists, it seems like the first lists (A-C) should overlap more with the second lists (D-F); I suggest instead to remove hypotheticals/unknowns, at least from the second lists (though Ln 244 suggests that unannotated genes were removed, so I am not sure how those were defined, considering that unknowns are still showing up here?). Panel E: How is phycoerythrin not considered at least somewhat photosynthesis-related?

Figure 2: Same comments about color. I do not understand what the clades are in D-F; please explain these labels in the figure caption. The figures should be stand-alone without requiring consultation of the main manuscript text to understand labels.

Figure 3: This is a nice figure aesthetically, but I wonder how well it actually represents meaningful data. The color-coding by the maximum peak would seem to artificially select for a visual trend. For comparative purposes, could the authors prepare the same figure for all three lakes, either as three panels here or with the other two panels in the supplement? Otherwise, it seems like there is the potential for cherry-picking the data that look nice. It could also be useful to see this as an all-genes comparison of the top x (100?) most highly expressed genes across the dataset in the supplement. I suspect that most of those genes will not show diel trends, which is fine, but that will help to put the genes that do show patterns in better context.

Reviewer: 2

Comments to the Author

This is an amazing dataset. But I couldn't find important information about study design and data analysis that would have allowed me to understand the analyses better. I may have missed some of this information, and I apologize if that's the case.

The two critical issues for me are:

1) The authors indicate that they used internal standards in the methods, but never mention when they are used in data analysis presented in the paper and when they aren't. Because the diel cycling of gene expression is a major focus of this paper, it is important to know when the data are actual counts per volume of lake water and when they are percentages. If the latter, the genes that appear to have higher night time expression may simply arise from the large change in photosynthesis transcripts. The units for Fig. 3, in particular, aren't explained.

2) The analysis yields no real insights. The major points seem to be that photosynthesis occurs in the day and heterotrophic bacteria take up sugars made by phytoplankton (Line 32 and 351, particularly). Since this is one of the largest metatranscriptome analyses yet done, and the study design is really nice, there must be some substantive insights to be gleaned from the data. Line 35 in the abstract seems to say that this paper discovered that photosynthesis occurs at different times during a diel cycle. I'm sure there is more nuance to the findings, but they do not come through in the manuscript.

Major comments:

Line 126: How many replicates for each lake and each time point? This info is particularly necessary to interpret the section on within versus between sample diversity.

Line 235. This section and Figure 1 are uninformative. A laundry list of highly expressed genes does not do justice to this novel dataset.

Line 305 paragaraph: Are the genes present in the genomes but not expressed, or they are missing from the genomes? For example, it's stated that the Actinobacteria only have R/S/M transporters differentially expressed; do they have ribose transporters that they aren't expressed differentially, or do they not have ribose transporters? The first option indicates differential transport by the Actinos but the second indicates genetic capability differences between taxonomic groups. Several statements in this paragraph could be interpreted either way. Also see Line 370 for another section where it's not clear if this is due to differences in gene expression or differences in genome content.

Line 342: I didn't see evidence for 'generalizable interactions'. Upregulation of photosynthesis during the day and phytoplankton supporting heterotrophic bacteria are already generalized and can't be claimed as a new finding here. Also, Line 362: bacterial uptake of sugars released from phytoplankton has been known for many decades.

Line 365: N=1 for each lake type, so this isn't the right dataset to address differences across major categories of lakes. The distinctions between these particular lakes might hold up across the classes if more data were available, but multiple types of each class would be needed to draw these conclusions. I don't mean to minimize the effort it took to get these data, just that this is not a hypothesis that can be robustly addressed.

Line 417: This study did not measure algal exudates. That might be a hypothesis to explain the diel transcription patterns, but this sentence goes beyond that.

Line 293: Why did you decide to lump all the day and night samples together for the statistical analyses? Since you have a high coverage, time-resolved dataset, why not to take advantage of this? Previous work (Ottesen and Aylward, for example) have found shifts in timing of gene expression during the day that have been really interesting. Perhaps novel findings will emerge if you change this to a full time-series analysis.

Figure 4 confused me. Why are the categories in different places for each lake? For example, if photosynthesis was in the upper left for all three, it would be easy to compare among the lakes. Some gene categories may be analyzed for only two lakes, but they could still be lined up with a blank region where no analysis was done (presumably because that gene category had no significant differences in that lake?). Did you try plotting these as total reads? It would be interesting to see how the number of transcripts being made by a taxonomic group changes between night and day.

Picky things:

Line 63: The term phototroph typically includes photoautotrophs and photoheterotrophs. There are some of both in this dataset. Defining phototrophs for this paper as just the photosynthetic microbes is confusing. Where do the photoheterotrophs fit in?

Line 98: 80% release of carbon as DOC is super high. I don't doubt that someone reported it somewhere, but the general consensus in the literature is ~20%. 99% doesn't even make sense, since there is nothing left for the phytoplankton for cell maintenance or division.

Line 257: Confusing wording; the genes of these taxonomic groups were expressed, not the taxonomic groups themselves.

Line 292: This is a key section for which the units on graphs should be make clear: are these percents of a pool, or transcript numbers normalized to volume filtered, or something else?

Line 321: 15% of the ROS genes, or 15% of all genes?