# Metatranscriptomics reveals interactions between phototrophs and heterotrophs in freshwater

Alexandra M. Linz1, Frank O. Aylward2, Stefan Bertilsson3, Katherine D. McMahon1,4

1Department of Bacteriology, University of Wisconsin–Madison, 2Department of Biological Sciences, Virginia Tech, 3Department of Ecology and Genetics, Limnology and Science for Life Laboratory, Uppsala University, Sweden, 4Department of Civil and Environmental Engineering, University of Wisconsin–Madison

## Abstract

## Introduction

Many of the core ecosystem functions in freshwater lakes, such as primary production and organic matter recycling, are largely driven by microbial communities. While the number of biochemical reactions performed by each individual cell is miniscule, the collective action of all these cells forms a dynamic, interconnected community whose emergent functions can impact ecosystem-level processes. Previous process rate measurements in a wide range of freshwater ecosystems indicate that day/night cycles drive photosynthesis, respiration, and dissolved organic matter (DOM) concentrations, suggesting metabolic interactions between the phototrophic and heterotrophic microbial communities in freshwater (Solomon et al. 2013; Kaplan and Bott 1989; Bertilsson and Jones 2003). We hypothesized that these diel trends would also be reflected in gene expression and sought to propose mechanisms of community interaction based on the timing of gene expression across the community. Therefore, we produced three two-day time series of metatranscriptomic sequencing data from three lakes with contrasting characteristics. We hypothesized that diel trends in gene expression occur in both phototrophic and heterotrophic microbes due to both the direct impacts of sunlight (such as driving photosynthesis and exposure to reactive oxygen species) and its indirect effects (such as metabolite exchange), regardless of lake type.

Previous metatranscriptomic work in marine and freshwater systems has highlighted potential links between phototrophic and heterotrophic microbes. One metatranscriptomic study in a phosphorus-limited mountain lake found differential gene expression between day and night in both phototrophs and heterotrophs, particularly in energy acquisition pathways and pyrophosphatase (Vila-Costa et al. 2013). Another study in marine systems also observed enriched expression of energy acquisition pathways during the day and higher expression of biosynthesis and housekeeping pathways at night (Poretsky et al. 2009). Strong diel patterns in both heterotrophic and phototrophic gene expression followed by a cascade of heterotrophic gene expression have also been observed in marine systems (Ottesen et al. 2014). A trend of peak activity in heterotrophic bacterioplankton following that of photoautotrophs was consistent in two different regions of the Pacific Ocean, potentially indicating strong metabolic connections between these two groups in disparate marine microbial communities (Aylward et al. 2015). These studies suggest that diel trends may be a universal characteristic of transcription-level response in aquatic microbial communities.

Other research also suggests strong connections between phototrophic and heterotrophic community members in aquatic ecosystems, and these connections may be driven by light. Co-cultures of phototrophic algae and heterotrophic bacteria can be stable, indicating cooperative interactions, although competition or predation have also been observed in the laboratory (Cole 1982; Pernthaler et al. 2001; Posch et al. 1999). In both marine and freshwater ecosystems, the community compositions of phototrophs and heterotrophs are inextricably linked (Paver et al. 2015, 2013). Perturbations in one portion of the community quickly ripple through the rest (Kent et al. 2006; Šimek et al. 2002). One potential mechanism that can explain these trends is DOM release by phototrophs, either through predation, cell detritus, leakage, or active transport. This DOM is composed of low molecular weight compounds, such as sugars, amino acids, carboxylic acids, and alditols (Maršálek and Rojíčková 1996; Hellebust 1965). In marine ecosystems, the percentage of photosynthetic carbon release extracellularly has been reported to be up to 80%, while percentages as high as 99% have been reported in freshwater (Bertilsson and Jones 2003). Although the factors leading to this DOM release are not fully understood, this DOM likely supports a substantial portion of the heterotrophic community. In fact, some ubiquitous freshwater bacteria, such as *Limnohabitans,* appear to specialize in algal-derived carbon uptake (Simek et al. 2011).

Sunlight itself may be a common factor driving heterotrophic and phototrophic gene expression in aquatic ecosystems. Photodegradation of complex DOM is another potential mechanism that could drive diel trends in heterotrophs by making recalcitrant DOM more labile (Bertilsson and Tranvik 2000; Jorgenson et al. 1998). The presence of opsins, extensively documented in both freshwater and marine heterotrophs, may also lead to cycles of diel gene expression in organisms predicted to be primarily heterotrophic (Pinhassi et al. 2016; Atamna-Ismaeel et al. 2008). Even without opsins, some freshwater microbes such as Actinobacteria may sense light in order to optimally time uptake and catabolism of organic substrates (Maresca et al. 2019). Finally, sunlight causes oxidative stress via UV irradiation and as a byproduct of photosynthesis (Sommaruga et al. 1997). Heterotrophic organisms may modulate their gene expression to avoid this stress, therefore introducing diel trends into these organisms.

To identify both biotic and abiotic trends driving community-level diel gene expression, we sequenced metatranscriptomes from the epilimnia of three freshwater lakes representing oligotrophic, eutrophic, and dystrophic (humic) lake types. These metatranscriptomes form a 2-day time series for each lake, with samples collected every four hours. We additionally collected sequenced metagenomes and single amplifed genomes (SAGs) from each lake to produce highly specific references, allowing us to obtain higher quality annotation and classification information than through read-based annotations. We observed diel trends in both phototrophic and heterotrophic microbial community members and were able to propose both biotic and abiotic mechanisms for these trends based on gene expression. Because this study is among the largest metatranscriptomic sequencing efforts to date, we discuss the biological vs. technical variability observed in this dataset to add to our knowledge of variability in environmental metatranscriptomics and to inform the future study designs (Tsementzi et al. 2014). Although different taxa and genes were expressed in the three lakes studied, we found consistent diel trends in each.

## Methods

### Study design and in situ measurements

Three lakes in Wisconsin, USA, were chosen for this study based on their different characteristics and trophic status: oligotrophic (Sparkling Lake, SL), eutrophic (Lake Mendota, LM), and humic (Trout Bog Lake, TB) (Table 1).LM is located in Madison, WI, USA, while TB and SL are located in Boulder Junction, WI, USA, approximately 350 km north of Madison. These lakes were chosen because they are core sites of the North Temperate Lakes - Long Term Ecological Research (NTL-LTER) program. They have a rich context of historical environmental data and automated sensor platforms were deployed at all three sites at the time of sampling. Previous studies of the resident microbial communities have also been performed in all three of these lakes, providing reference genomes specific to each of these lakes (Linz et al. 2018; Ghylin et al. 2014; Bendall et al. 2016).

Hereafter, we provide brief summaries of our methods; full protocols are available in Supplemental Document X. The epilimnion (top thermal layer) of each lake was sampled twelve times at four-hour intervals, starting at 5:00AM and continuing until 1:00AM, 44 hours later. The lakes were sampled in July 2016 within a two week time period to minimize seasonal changes. Due to the difference in latitude, the day length at SL and TB was slightly longer than at LM. Half an hour prior to each timepoint, an instrumented sonde (Hydrolab DS5X, OTT Hydromet, Kempten, Germany) equipped with sensors for temperature, dissolved oxygen concentrations, pH, conductivity, and turbidity was used to collect measurements from the top ten meters of the water column (in TB, which is only eight meters deep, the whole water column was measured). Photosynthetically active radiation (PAR) was also measured at this time using a PAR meter (Li-Cor, Lincoln, NE, USA). PAR readings were obtained every 0.5 meter, until light extinction or six meters. Secchi depth was measured once per lake during the time series.

At each timepoint, we collected an integrated water sample of the epilimnion. The sampling depth was chosen based on the location of the thermocline on the day prior to initiation of the two-day time series of the respective lake. The collection depth was kept constant throughout the sampling period. All collection tools and bottles were washed with ambient epilimnion water prior to each timepoint. To collect RNA, water from the integrated epilimnion sample was immediately pumped through 0.22-m polyethylene filters (Pall, Port Washington, NY, USA) with a cheesecloth pre-filtration. This process occurred in the field using a Masterflex E/S portable sampler (Cole-Parmer, Vernon Hills, IL, USA). Each sample was filtered for the same amount of time based on the rate of filter clogging (consistently 2.5min for Sparkling and Trout, 1.25-3min for Mendota). Filters were placed in 2-mL plastic cryovials (Phenix Research, Candler, NC, USA) and immediately flash frozen in liquid nitrogen in the field. Filters were stored at -80oC after collection, and samples from Trout Bog and Sparkling samples spent four hours on dry ice during transport to Madison.

After filtering for RNA, additional samples were collected for lab-based measurements of environmental variables from the same epilimnion sample. Samples for total and dissolved nitrogen and phosphorus concentrations were collected in 60-mL HDPE bottles (Nalgene, Rochester, NY, USA). One liter of unfiltered water for chlorophyll measurement was collected in a black, opaque glass bottle and filtered on shore (approximately 30 minutes after collection) through a 0.3m nitrocellulose Whatman filter (GE Healthcare Life Sciences, Pittsburgh, PA, USA) and flash frozen in liquid nitrogen. Unfiltered water was collected for protein synthesis assays using 14C-leucine (Chin-Leo and Kirchman 1988). DNA samples for metagenomic sequencing were collected at one timepoint from each lake by filtering 250 mL of the integrated epilimnion water sample through the same type of 0.22-m filters used for RNA collection. Cells were preserved for single amplified genome (SAG) sequencing by mixing 2-mL of the integrated epilimnion sample with 100-uL of a glycerol-TE buffer. Both the DNA and single cell preservation samples were flash frozen in liquid nitrogen and stored at -80oC until processing.

### RNA extraction

Within 2-3 weeks of collection, RNA was extracted from the filters in a single batch operation. Filters were exposed to a lysis solution containing EDTA and SDS and incubated at 65 oC. Filters were then physically destroyed using FastDNA Spin Kit reagents and bead-beating protocol (MP Biomedicals, Santa Ana, CA, USA). TRIzol (Thermo-Fisher, Waltham, MA, USA) was added to the filters before physical disruption. An internal standard - an *in vitro* transcription of the cloning plasmid pFN18A - was prepared prior to extraction as in Satinsky, et al. 2013 and added to samples after beadbeating (Satinsky et al. 2013). Phenol:chloroform was used to isolate RNA from the lysate. After phenol:chloroform purifications, the RNA was precipitated in ethanol, pelleted, and resuspended in nuclease-free water. The RNA was further purified using an RNeasy kit (QIAGEN, Hilden, Germany) with an on-column DNAse digestion. All samples were quantified using a Qubit fluorimeter (Thermo-Fisher) and stored at -80oC until sequencing. A subset of the samples were further tested on a BioAnalyzer to confirm that the RNA was of sufficient quality for sequencing (Agilent Genomics, Santa Clara, CA, USA).

### Additional lab-based measurements

Chlorophyll was extracted from the triplicate filters using methanol following NTL-LTER protocols (https://lter.limnology.wisc.edu/protocol/spectrophotometric-chlorophyll-measurement). Bacterial production assays were conducted at each timepoint using 14C-leucine (Chin-Leo and Kirchman 1988) and frozen after stopping incubations with 100% trichloroacetic acid. Approximately one month after sample collection, production assay samples were thawed, the protocol was completed, and samples were assayed for 14C radioactivity. Total and dissolved nitrogen and phosphorus were measured via colormetric HPLC following LTER protocols (<https://lter.limnology.wisc.edu/protocol/total-phosphorus-and-total-nitrogen>) using filtered and unfiltered aliquots of water collected in 60-mL HDPE bottles and stored at -20oC for approx. 4 months. DNA filters underwent a phenol/chloroform extraction using the same lysis method as the RNA extraction protocol. Four additional DNA samples collected from Sparkling Lake in 2009 were extracted and sent for sequencing to serve as additional references for this lake.

### Sequencing

Three samples from each timepoint were sequenced by the Department of Energy Joint Genome Institute (JGI). Once received, ribosomal RNA was depleted from the RNA samples, and they were converted to cDNA. The resulting cDNA was sequenced using Illumina HiSeq 2500-1TB (Illumina, San Diego, CA, USA). Metatranscriptomic reads were quality filtered by the JGI. Metatranscriptomic reads were assembled by the JGI using MetaHit (Nielsen et al. 2014). DNA samples for metagenomics were also sequenced on an Illumina HiSeq platform. Metagenomic reads from each individual sample were quality filtered and assembled using MetaHit, at the JGI. DNA samples for 16S rRNA ribosomal gene amplicon sequencing were sequenced on an Illumina MiSeq platform. The resulting reads were filtered using BBDuk and reads mapping to human, mouse, cat, and dog genomes with BBMap were removed (Bushnell et al. 2014).

Cells for single amplified genomes were sorted, identified using 16S rRNA gene amplicon sequencing, and sequenced using the JGI’s standard single amplified genome (SAG) protocols. Cells for SAG sequencing were chosen with a preference for Sparkling, the least well-represented lake in our pre-existing reference genome collection. An Illumina shotgun library was constructed from each single cell and sequenced on the Illumina NextSeq platform. Sequencing reads were filtered using BBTools and assembled into SAGs using SPAdes (Bankevich et al. n.d.). Unscreened SAGs were used as references to retain any unusual DNA sequences in the genome.

### Bioinformatics pipeline

In total, our sequencing effort produced approximately 2 terabytes of data. We designed our bioinformatics pipeline to run in a high-throughput computing system in order to handle this data volume. Code and pipeline details can be found at https://github.com/McMahonLab/geodes.

Ribosomal RNA reads, which still comprised approximately 50% of metatranscriptomic reads despite depletion prior to sequencing, were first removed using SortMeRNA (Kopylova, Noé, and Touzet 2012). Assembled metagenomic contigs from this study, SAGs from this study, SAGs and MAGs from previous McMahon Lab time series sequencing on these lakes (Linz et al. 2018; Ghylin et al. 2014; Garcia et al. 2018), and 5 freshwater algal genomes from NCBI RefSeq (Pruitt and Maglott 2001), representing each available algal genus, were used to build a nonredundant, highly specific database for subsequent mapping of metatranscriptomic reads (supp table of IMG Genome IDs used in database). This approach provides better functional prediction than annotating each individual read. After formatting each type of genome or contig’s fastq and gff files, coding regions were extracted and clustered at 97% ID using CD-HIT (Huang et al. 2010). Metatranscriptomic reads were mapped to this database with a 90% ID cutoff using BBMap and requiring at least 75% overlap with a gene feature. Mapped reads were tabulated using FeatureCounts (Liao, Smyth, and Shi 2014). Individual metagenome assemblies were binned using Metabat (Kang et al. 2015) and checked for completeness and contamination using CheckM (Parks et al. 2015). Bins and unbinned contigs from the metagenome assemblies were classified by taking the consensus taxonomy of the best hit in the Integrated Microbial Genomes database (Markowitz et al. 2012) for each coding region on a contig/bin (Stevens, unpublished).

Addition of an internal RNA standard allowed for both normalization of expressed reads and assessment of extraction success. Samples with either too few counts of the internal standard (less than 50) or orders of magnitude higher expression of all genes after normalization when compared to replicates were discarded. After these quality control measures, 32 samples remained from Sparkling, 30 from Mendota, and 21 from Trout Bog. Many samples from day two in the Trout Bog time series failed to meet to quality control standards.

### Statistics

The statistical software R was used for further analysis (R Core Team, 2018). To reduce noise in the dataset, the top 20,000 expressed genes in each lake were retained for further differential expression analysis. From this subset, marker genes for metabolic processes were selected and aggregated by pathway. The summed expression of each pathway/process was input into DESeq to test differential expression (cite). Using the internal standard to determine normalization size factors, we converted read counts to units of transcripts per liter. Therefore, these results are semi-quantitative (keeping in mind the inherent limitations and biases of metatranscriptomics from water collection through sequencing) (Tsementzi et al. 2014). In addition to normalizing by the internal standard, samples were still normalized using a negative binomial distribution using DESeq to control for compositional bias before testing differential expression (Anders and Huber 2010). RAIN was used to detect cyclic trends in gene expression (cite). Based on our PAR measurements (supp), day timepoints were considered to be 9AM, 1PM, and 5PM, while night timepoints were considered to be 9PM, 1AM, and 5AM. Results were plotted using the R packages ggplot2 (Wickham, 2009) and cowplot (Wilke, 2017). All code is available at <github link>.

## Results

### What genes were expressed?

As an initial comparison between our study sites, we first asked which genes were most expressed in each lake (Table 2). Photosynthesis related genes, particularly those relating to photosystem II P680, were highly expressed in all three lakes. Genes encoding ribulose-1,5-bisphosphate carboxylase (RuBisCO), the key enzyme in carbon fixation via the Calvin-Benson-Bassham (CBB) pathway, were among the most highly expressed genes in Lake Mendota and Trout Bog. These genes were most frequently derived from *Cyanobacteria.*

Because of the high expressions of genes related to phototrophy in all sites, we also ran this analysis excluding genes associated with phototrophy (Table 3). This showed that housekeeping genes such as RNA polymerase, chaperonin, and translation elongation factors were commonly expressed in all lakes. Many of the most highly expressed heterotrophic genes in Lake Mendota were classified as belonging to acI Actinobacteria, including a sugar transporter. In Trout Bog, *Verrucomicrobia* and *Armatimonadetes* (formerly candidate phylum OP10) contributed some of the top expressed genes, while in Sparkling Lake, a chaperonin expressed by *Deltaproteobacteria* was the most highly expressed heterotrophic gene. Cytochrome subunits, essential components of respiratory metabolisms, were highly expressed in all lakes, ranking in the top 10 in Trout Bog and Sparkling Lake, and in the top 25 in Mendota.

*Which taxa were expressing genes?*

We next aggregated expressed genes by phylum-level classifications to compare the most expressed taxa to the most abundant taxa based on metagenomic data (Figure 1). The same reference database was used for mapping metatranscriptomic and metagenomic data, making such comparisons possible. No positive trend between expression and abundance was observed. Eukaryotic algae, including *Heterokonta, Streptophyta, and Cryptophyta,* were highly expressed in all three lakes, with more types of algae observed in Trout Bog. *Cyanobacteria* were highly expressed in all three lakes, while viruses were also present, but expressing at low levels in all sites. The only abundant Archaeal phylum observed in the metatranscriptomes was *Crenarchaeota*. One phylum, *Chloroflexi*, had orders of magnitude higher expression and abundance than other phyla in Lake Mendota. This phylum is likely an misclassified - genes with this classification were almost exclusively derived from a single, low quality MAG.

### Trends in environmental variables

We examined a suite of potentially relevant environmental variables to compare trends in these to those observed in gene expression, expecting that several of these trends would be diel. Parameters that reflect the boundaries between layers within the water column, such as dissolved oxygen, temperature, pH, and conductivity, were strongly diel in Lake Mendota, but less so in Sparkling Lake and Trout Bog (supplemental). Chlorophyll concentrations, often used as an indicator of primary production, were diel in Trout Bog, but not in the other two sites. Bacterial production, measured via 14C-leucine incorporation, showed dynamics over the two day time series in all three lakes, although the trends were not diel (supplemental). No trends were observed in total and dissolved nitrogen or phosphorus concentrations.

### Assessing variability in freshwater metatranscriptomes

One of the goals of this experiment was to determine the amount of variability in freshwater gene expression to inform future metatranscriptomic experiments. We used the coefficient of variation (CoV), the ratio of standard deviation to average expression (%), to compare the amount of variability within replicate samples to the variation observed across different timepoints (Figure 2). Higher CoVs were observed across samples than within replicates. Still, the upper limit for CoV within replicates approached 200%.

### Gene expression in day vs. night

To test differential expression in day vs. night, we aggregated timepoints by day (9AM, 1PM, and 5PM) or night (9PM, 1AM, and 5AM). To reduce the number of comparisons performed, this analysis was performed on the top 20,000 most abundantly expressed genes in each lake. We identified genes with significant differential expression in day vs. night and tested for significant differences in the number of reads assigned to genes in functional categories. We also used RAIN to reveal any cyclic trends with 12-hour periods among genes already pre-screened as having differential expression in day vs. night.

Genes related to photosynthesis were significantly more expressed in day vs. night and contained the highest numbers of cyclic genes in all lakes (Tables 2, 3, 4). In Mendota, this expression was largely derived from *Cyanobacteria,* while photosynthesis-related gene expression in Trout Bog and Sparkling was derived from a mix of *Cyanobacteria, Eukaryota,* and unclassified groups*.* Expression of genes encoding the key carbon fixation enzyme RuBisCO was only significantly different in day vs. night in Trout Bog, where it was 7-fold higher during the day and the associated gene clusters were largely taxonomically unclassified.

Genes related to sugar transport were often significantly more expressed at night in all three lakes. Specifically, genes annotated as general sugar transporters, ribose transporters, and raffinose/stachyose/melibiose transporters were significantly more expressed at night than during the day in Mendota (Figure 3). General sugar transporters were expressed by *Actinobacteria* (particularly acI-B1 and acTH1-A1)*, Cyanobacteria* (predominantly *Synechococcaceae* and *Microcystis*)*,* and *Bacteroidetes,* with a lower proportion of reads derived from *Cyanobacteria* at night compared to daytime expression. Ribose transporters and raffinose/stachyose/melibiose transporters were mostly classified as *Actinobacteria* and *Bacteroidetes,* with little difference in profiles between day and night. In Trout Bog, genes annotated as transporters for general sugars, ribose, and xylose were significantly more expressed at night. *Actinobacteria* (particularly acI-B)contributed the majority of expressed reads for all three types of sugar transporters, with *Alphaproteobacteria* identified in xylose and general sugar transport, and *Cyanobacteria* and *Armatimonadetes* also contributing to general sugar transport. The only significant differentially expressed sugar transport group in Sparkling was raffinose/stachyose/melibiose transport in *Actinobacteria*, although several other types were near the significance threshold.

Reactive oxygen species (ROS) defense is a critical function for microbes during the day. As expected, genes related to ROS defense were significantly more expressed in day vs night in all three lakes, with roughly 15% of genes identified as cyclic in Mendota and Sparkling. Phyla expressing ROS defense-related genes in Mendota included *Cyanobacteria, Deltaproteobacteria, Planctomycetes, Verrucomicrobia, Betaproteobacteria, Bacteroidetes,* and *Alphaproteobacteria.* In Trout Bog, ROS defense-related reads were assigned to *Actinobacteria, Alphaproteobacteria, Armatimonadetes, Bacteroidetes, Betaproteobacteria, Cyanobacteria, Eukaryota, Gammaproteobacteria,* and *Verrucomicrobia,* with roughly a 3rd of reads mapping to unclassified gene clusters. Fewer major phyla were expressing ROS defense genes in Sparkling, where the majority were contributed by *Betaproteobacteria,* followed by *Actinobacteria* and *Eukaryota,* and a number of low-expression phyla contributing up to half of the reads. A greater proportion of reads were assigned to *Eukaryota* at night vs. day.

Several functional gene categories differed in significance between lakes beyond those already mentioned. Genes related to rhodopsins, classified as *Actinobacteria* (largely acI-B1)or *Bacteroidetes,* were significantly more expressed in day in Mendota, as were genes related to proteases (*Cyanobacteria, Betaproteobacteria*, *Gammaproteobacteria, Bacteroidetes,* and *Actinobacteria,* with a higher proportion of *Cyanobacteria* in day). As previously discussed, genes associated with xylose transport and RuBisCO were significantly differentially expressed in Trout Bog. Amino acid transport genes were more expressed at night in Trout Bog and these were classified as belonging to *Actinobacteria, Alphaproteobacteria, Armatimonadetes,* and *Betaproteobacteria*. Fewer functional groups were differentially expressed in Sparkling compared to the other two lakes, with no groups found to be significant only in Sparkling.

## Discussion

In this study, we sought to identify biotic and abiotic factors driving diel gene expression across multiple lake types. Using metatranscriptomic time series, we were able to detect genes that were differentially expressed in day vs. night and identify those that showed cyclic trends in the time series. The functional annotations of those genes allow us to hypothesize which forces drive gene expression in freshwater microbial communities.

The balance of primary productivity and overall respiration rates is of great interest to limnologists seeking to create carbon budgets for freshwater lakes, we focused on these processes. Indeed, previous work using high-frequency dissolved oxygen measurements has linked photosynthesis and respiration to diel cycles (Solomon et al. 2013), leading us to hypothesize that genes related to these processes would also show diel trends. In all three lakes, genes related to photosynthesis were more highly expressed in day and most often cyclic, compared to other functional categories. Oxygenic photosynthesis performed by *Cyanobacteria* or eukaryotic algae was most common, consistent with the oxygenated conditions in epilimnia. In ecosystem level analyses, photosynthesis and carbon fixation are often considered to be linked as primary production. However, we only saw differential expression of RuBisCO genes in Trout Bog. Still, expression does not necessarily correlate to protein abundance or function (cite), indicating only that transcriptional regulation of RuBisCO is not diel in Mendota or Sparkling.

Respiration is a broader category that could encompass the degradation of many carbon substrates. To narrow down the possible compounds being respired, we focused on genes related to carbon transport, as transporter expression has previously been used in marine systems to predict substrate use (Ottesen et al. 2013). Again, higher expression does not necessarily correlate to increased function, but it does indicate a signal to upregulate those functions. For transporters, availability of their target substrate is a common trigger for increased expression. In all three lakes studied, we found significantly higher expression of genes related to sugar transport at night compared to day. Phytoplankton are known to exude sugars (cite), suggesting a strong linkage between phototrophs and heterotrophs mediated by such sugars.

There are two non-exclusive hypotheses as to why we observed diel trends in genes encoding sugar transport. One is biotic in origin – if these sugars are indeed algal exudates, they may be produced during the day and released at night. Although the diel release of sugars from freshwater phytoplankton has not been observed, day/night partitioning of photosynthesis and sugar production occurs within plant and algae cells (cite). This diel trend may extend beyond the single cell to community-level interactions. The other hypothesis is that sunlight-generated oxidative stress prevents many freshwater microbes from consuming sugar during the day, even if it is available. We observed differential expression in genes related to reactive oxygen species defense such as catalases and peroxidases, with higher expression observed during the day in all three lakes. Although sunlight can cause DOM degradation, we did not observe differential expression of genes encoding transporters for typical photodegradation products, such as glycolate or carboxylic acids.

There is evidence from marine microbial communities suggesting that carbon released by phototrophic community members influences heterotrophic community composition in a way that improves phototroph fitness. The marine phototroph *Prochlorococcus* likely exudates carbon to maintain redox balance, as it generates more reducing power via photosynthesis than it can use (Bertilsson et al. 2005). However, a frequently observed adaptation to excess reducing power is to downregulate photosynthesis electron flux; this is not observed in *Prochlorococcus* and suggests alternative reasons for its release of carbon (Braakman, Follows, and Chisholm 2017)*.* In marine communities, heterotrophic bacterioplankton are highly dependent upon these *Prochlorococcus* exudates and likely perform a critical community function in return, such as the detoxification of hydrogen peroxide or free radicals. Again using *Prochlorococcus* as an example, this phototroph has lost its genes for reactive oxygen species defense and depends on the heterotrophic community for this function (Ma et al. 2018; Morris et al. 2016). In coral reefs, algal exudates can dramatically shift bacterial community composition, potentially providing algae with a competitive advantage over coral by selecting for coral pathogens in the heterotrophic community (Nelson et al. 2013).

It is therefore reasonable to hypothesize that freshwater photoautotrophs may be releasing carbohydrates to shape the heterotrophic community for their own benefit. Perhaps heterotrophs produce compounds that are beneficial for phototrophs, such as peroxidases or catalases, vitamins, antimicrobial peptides and antibiotics, or inorganic nutrients (all of which were expressed in our metatranscriptomic dataset). The origin of metabolic exchanges that lead to co-dependencies has been postulated to be an important driver of evolution in aquatic communities (The so-called “Black Queen Hypothesis” (Morriss et al., mBio, 2012)). It is intriguing to note that the dependency between phototrophs and heterotrophs and the diel partitioning of carbon fixation and respiration would be analogous to the organization and functioning of chloroplasts and mitochondria in plant cells. However, further experimentation is needed to confirm these hypotheses.

Here, we present a comparative metatranscriptomic analysis which demonstrates similar diel trends in photosynthesis, reactive oxygen species defense, and sugar transport in three different types of lakes, suggesting that these trends may be a general property of freshwater microbial communities. We outline both biotic (algal exudates) and abiotic (oxidative stress) that may be driving community-level diel trends in freshwater. Whether all of these microbes are responding to the same day-night stimulus or whether community interactions drive these diel trends remains to be determined. We also provide lists of the most expressed genes and phyla, as well as an assessment of the variability observed in our dataset, to guide future metatranscriptomic studies. All data and code are publicly available at <github link>. Our detection of diel trends in freshwater is the first of many analyses that can be performed on this extensive set of metatranscriptomic time series.

I think you need a paragraph discussing differences you saw across lakes. Of course we have to temper our conclusions because we only have one of each lake, but the differences you point out in the results deserve some interpretation.

## References

Anders, Simon, and Wolfgang Huber. 2010. “Differential Expression Analysis for Sequence Count Data.” *Genome Biology* 11 (10): R106. https://doi.org/10.1186/gb-2010-11-10-r106.

Atamna-Ismaeel, Nof, Gazalah Sabehi, Itai Sharon, Karl-Paul Witzel, Matthias Labrenz, Klaus Jürgens, Tamar Barkay, Maayke Stomp, Jef Huisman, and Oded Beja. 2008. “Widespread Distribution of Proteorhodopsins in Freshwater and Brackish Ecosystems.” *The ISME Journal* 2 (6): 656–62. https://doi.org/10.1038/ismej.2008.27.

Aylward, Frank O., John M. Eppley, Jason M. Smith, Francisco P. Chavez, Christopher A. Scholin, and Edward F. DeLong. 2015. “Microbial Community Transcriptional Networks Are Conserved in Three Domains at Ocean Basin Scales.” *Proceedings of the National Academy of Sciences* 112 (17): 5443–48. https://doi.org/10.1073/PNAS.1502883112.

Bankevich, Anton, Sergey Nurk, Dmitry Antipov, Alexey A Gurevich, Mikhail Dvorkin, Alexander S Kulikov, Valery M Lesin, et al. n.d. “Original Articles SPAdes: A New Genome Assembly Algorithm and Its Applications to Single-Cell Sequencing.” Accessed December 14, 2018. https://doi.org/10.1089/cmb.2012.0021.

Bendall, Matthew L, Sarah LR Stevens, Leong-Keat Chan, Stephanie Malfatti, Patrick Schwientek, Julien Tremblay, Wendy Schackwitz, et al. 2016. “Genome-Wide Selective Sweeps and Gene-Specific Sweeps in Natural Bacterial Populations.” *The ISME Journal* 10 (7): 1589–1601. https://doi.org/10.1038/ismej.2015.241.

Bertilsson, Stefan, Olof Berglund, Michael J Pullin, and Sallie W Chisholm. 2005. “Release of Dissolved Organic Matter by Prochlorococcus.” *Vie et Milieu* 55 (3–4): 225–32. https://www.researchgate.net/publication/252613206.

Bertilsson, Stefan, and Jeremy B. Jones. 2003. “Supply of Dissolved Organic Matter to Aquatic Ecosystems: Autochthonous Sources.” *Aquatic Ecosystems*, January, 3–24. https://doi.org/10.1016/B978-012256371-3/50002-0.

Bertilsson, Stefan, and Lars J. Tranvik. 2000. “Photochemical Transformation of Dissolved Organic Matter in Lakes.” *Limnology and Oceanography* 45 (4): 753–62. https://doi.org/10.4319/lo.2000.45.4.0753.

Braakman, Rogier, Michael J Follows, and Sallie W Chisholm. 2017. “Metabolic Evolution and the Self-Organization of Ecosystems.” *Proceedings of the National Academy of Sciences of the United States of America* 114 (15): E3091–3100. https://doi.org/10.1073/pnas.1619573114.

Bushnell, Brian, Rob Egan, Alex Copeland, Brian Foster, Alicia Clum, Hui Sun, Kurt Labutti, et al. 2014. “BBMap: A Fast, Accurate, Splice-Aware Aligner.” https://doi.org/10.1186/1471-2105-13-238.

Chin-Leo, G, and D L Kirchman. 1988. “Estimating Bacterial Production in Marine Waters from the Simultaneous Incorporation of Thymidine and Leucine.” *Applied and Environmental Microbiology* 54 (8): 1934–39. http://www.ncbi.nlm.nih.gov/pubmed/16347706.

Cole, Jonathan J. 1982. “INTERACTIONS BETWEEN BACTERIA AND ALGAE IN AQUATIC ECOSYSTEMS.” *Ann Rev. Ecol. Syst*. Vol. 13. www.annualreviews.org.

Garcia, Sarahi L., Sarah L. R. Stevens, Benjamin Crary, Manuel Martinez-Garcia, Ramunas Stepanauskas, Tanja Woyke, Susannah G. Tringe, et al. 2018. “Contrasting Patterns of Genome-Level Diversity across Distinct Co-Occurring Bacterial Populations.” *The ISME Journal* 12 (3): 742–55. https://doi.org/10.1038/s41396-017-0001-0.

Ghylin, Trevor W, Sarahi L Garcia, Francisco Moya, Ben O Oyserman, Patrick Schwientek, Katrina T Forest, James Mutschler, et al. 2014. “Comparative Single-Cell Genomics Reveals Potential Ecological Niches for the Freshwater AcI Actinobacteria Lineage.” *The ISME Journal* 8 (12): 2503–16. https://doi.org/10.1038/ismej.2014.135.

Hellebust, J. A. 1965. “EXCRETION OF SOME ORGANIC COMPOUNDS BY MARINE PHYTOPLANKTON1.” *Limnology and Oceanography* 10 (2): 192–206. https://doi.org/10.4319/lo.1965.10.2.0192.

Huang, Ying, Beifang Niu, Ying Gao, Limin Fu, and Weizhong Li. 2010. “CD-HIT Suite: A Web Server for Clustering and Comparing Biological Sequences.” *Bioinformatics* 26 (5): 680–82. https://doi.org/10.1093/bioinformatics/btq003.

Jorgenson, Niels OG, Lars J. Tranvik, Helene Edling, Wilhelm Graneli, and Mans Lindell. 1998. “Effects of Sunlight on Occurrence and Bacterial Turnover of Specific Carbon and Nitrogen Compounds in Lake Water.” *FEMS Microbiology Ecology* 25: 217–27.

Kang, Dongwan D, Jeff Froula, Rob Egan, and Zhong Wang. 2015. “MetaBAT, an Efficient Tool for Accurately Reconstructing Single Genomes from Complex Microbial Communities.” *PeerJ* 3: e1165. https://doi.org/10.7717/peerj.1165.

Kaplan, Louis A., and Thomas L. Bott. 1989. “Diel Fluctuations in Bacterial Activity on Streambed Substrata during Vernal Algal Blooms: Effects of Temperature, Water Chemistry, and Habitat.” *Limnology and Oceanography* 34 (4): 718–33. https://doi.org/10.4319/lo.1989.34.4.0718.

Kent, Angela D., Stuart E. Jones, George H. Lauster, James M. Graham, Ryan J. Newton, and Katherine D. McMahon. 2006. “Experimental Manipulations of Microbial Food Web Interactions in a Humic Lake: Shifting Biological Drivers of Bacterial Community Structure.” *Environmental Microbiology* 8 (8): 1448–59. https://doi.org/10.1111/j.1462-2920.2006.01039.x.

Kopylova, Evguenia, Laurent Noé, and Hélène Touzet. 2012. “SortMeRNA: Fast and Accurate Filtering of Ribosomal RNAs in Metatranscriptomic Data.” *Bioinformatics* 28 (24): 3211–17. https://doi.org/10.1093/bioinformatics/bts611.

Liao, Y., G. K. Smyth, and W. Shi. 2014. “FeatureCounts: An Efficient General Purpose Program for Assigning Sequence Reads to Genomic Features.” *Bioinformatics* 30 (7): 923–30. https://doi.org/10.1093/bioinformatics/btt656.

Linz, Alexandra M., Shaomei He, Sarah L.R. Stevens, Karthik Anantharaman, Robin R. Rohwer, Rex R. Malmstrom, Stefan Bertilsson, and Katherine D. McMahon. 2018. “Freshwater Carbon and Nutrient Cycles Revealed through Reconstructed Population Genomes.” *PeerJ* 6 (December): e6075. https://doi.org/10.7717/peerj.6075.

Ma, Lanying, Benjamin C Calfee, J Jeffrey Morris, Zackary I Johnson, and Erik R Zinser. 2018. “Degradation of Hydrogen Peroxide at the Ocean’s Surface: The Influence of the Microbial Community on the Realized Thermal Niche of Prochlorococcus.” *The ISME Journal* 12 (2): 473–84. https://doi.org/10.1038/ismej.2017.182.

Maresca, Julia A, Jessica L Keffer, Priscilla Hempel, Shawn W Polson, Olga Shevchenko, Jaysheel Bhavsar, Deborah Powell, Kelsey J Miller, Archana Singh, and Martin W Hahn. 2019. “Light Modulates the Physiology of Non-Phototrophic Actinobacteria.” *Journal of Bacteriology*, January, JB.00740-18. https://doi.org/10.1128/JB.00740-18.

Markowitz, Victor M., I. Min A. Chen, Krishna Palaniappan, Ken Chu, Ernest Szeto, Yuri Grechkin, Anna Ratner, et al. 2012. “IMG: The Integrated Microbial Genomes Database and Comparative Analysis System.” *Nucleic Acids Research* 40 (D1): 115–22. https://doi.org/10.1093/nar/gkr1044.

Maršálek, Blahoslav, and Renata Rojíčková. 1996. “Stress Factors Enhancing Production of Algal Exudates: A Potential Self-Protective Mechanism?” *Zeitschrift Für Naturforschung C* 51 (9–10): 646–50. https://doi.org/10.1515/znc-1996-9-1008.

Morris, J. Jeffrey, Zackary I. Johnson, Steven W. Wilhelm, and Erik R. Zinser. 2016. “Diel Regulation of Hydrogen Peroxide Defenses by Open Ocean Microbial Communities.” *Journal of Plankton Research* 38 (4): 1103–14. https://doi.org/10.1093/plankt/fbw016.

Nelson, Craig E, Stuart J Goldberg, Linda Wegley Kelly, Andreas F Haas, Jennifer E Smith, Forest Rohwer, and Craig A Carlson. 2013. “Coral and Macroalgal Exudates Vary in Neutral Sugar Composition and Differentially Enrich Reef Bacterioplankton Lineages.” *The ISME Journal* 7 (5): 962–79. https://doi.org/10.1038/ismej.2012.161.

Nielsen, H Bjørn, Mathieu Almeida, Agnieszka Sierakowska Juncker, Simon Rasmussen, Junhua Li, Shinichi Sunagawa, Damian R Plichta, et al. 2014. “Identification and Assembly of Genomes and Genetic Elements in Complex Metagenomic Samples without Using Reference Genomes.” *Nature Biotechnology* 32 (8): 822–28. https://doi.org/10.1038/nbt.2939.

Ottesen, Elizabeth A, Curtis R Young, John M Eppley, John P Ryan, Francisco P Chavez, Christopher A Scholin, and Edward F DeLong. 2013. “Pattern and Synchrony of Gene Expression among Sympatric Marine Microbial Populations.” *Proceedings of the National Academy of Sciences of the United States of America* 110 (6): E488-97. https://doi.org/10.1073/pnas.1222099110.

Ottesen, Elizabeth A, Curtis R Young, Scott M Gifford, John M Eppley, Roman Marin, Stephan C Schuster, Christopher A Scholin, and Edward F DeLong. 2014. “Ocean Microbes. Multispecies Diel Transcriptional Oscillations in Open Ocean Heterotrophic Bacterial Assemblages.” *Science (New York, N.Y.)* 345 (6193): 207–12. https://doi.org/10.1126/science.1252476.

Parks, Donovan H, Michael Imelfort, Connor T Skennerton, Philip Hugenholtz, and Gene W Tyson. 2015. “CheckM: Assessing the Quality of Microbial Genomes Recovered from Isolates, Single Cells, and Metagenomes.” *Genome Research* 25 (7): 1043–55. https://doi.org/10.1101/gr.186072.114.

Paver, Sara F., Kevin R. Hayek, Kelsey A. Gano, Jennie R. Fagen, Christopher T. Brown, Austin G. Davis-Richardson, David B. Crabb, et al. 2013. “Interactions between Specific Phytoplankton and Bacteria Affect Lake Bacterial Community Succession.” *Environmental Microbiology* 15 (9): 2489–2504. https://doi.org/10.1111/1462-2920.12131.

Paver, Sara F., Nicholas D. Youngblut, Rachel J. Whitaker, and Angela D. Kent. 2015. “Phytoplankton Succession Affects the Composition of *P* *Olynucleobacter* Subtypes in Humic Lakes.” *Environmental Microbiology* 17 (3): 816–28. https://doi.org/10.1111/1462-2920.12529.

Pernthaler, J, T Posch, K Simek, J Vrba, A Pernthaler, F O Glöckner, U Nübel, R Psenner, and R Amann. 2001. “Predator-Specific Enrichment of Actinobacteria from a Cosmopolitan Freshwater Clade in Mixed Continuous Culture.” *Applied and Environmental Microbiology* 67 (5): 2145–55. https://doi.org/10.1128/AEM.67.5.2145-2155.2001.

Pinhassi, Jarone, Edward F. DeLong, Oded Béjà, José M. González, and Carlos Pedrós-Alió. 2016. “Marine Bacterial and Archaeal Ion-Pumping Rhodopsins: Genetic Diversity, Physiology, and Ecology.” *Microbiol. Mol. Biol. Rev.* 80 (4): 929–54. https://doi.org/10.1128/MMBR.00003-16.

Poretsky, Rachel S., Ian Hewson, Shulei Sun, Andrew E. Allen, Jonathan P. Zehr, and Mary Ann Moran. 2009. “Comparative Day/Night Metatranscriptomic Analysis of Microbial Communities in the North Pacific Subtropical Gyre.” *Environmental Microbiology* 11 (6): 1358–75. https://doi.org/10.1111/j.1462-2920.2008.01863.x.

Posch, T, K Simek, J Vrba, J Pernthaler, J Nedoma, B Sattler, B Sonntag, and R Psenner. 1999. “Predator-Induced Changes of Bacterial Size-Structure and Productivity Studied on an Experimental Microbial Community.” *Aquatic Microbial Ecology* 18 (3): 235–46. https://doi.org/10.3354/ame018235.

Pruitt, K. D., and Donna R. Maglott. 2001. “RefSeq and LocusLink: NCBI Gene-Centered Resources.” *Nucleic Acids Research* 29 (1): 137–40. https://doi.org/10.1093/nar/29.1.137.

Satinsky, Brandon M., Scott M. Gifford, Byron C. Crump, and Mary Ann Moran. 2013. “Use of Internal Standards for Quantitative Metatranscriptome and Metagenome Analysis.” *Methods in Enzymology* 531 (January): 237–50. https://doi.org/10.1016/B978-0-12-407863-5.00012-5.

Simek, Karel, Vojtĕch Kasalický, Eliska Zapomĕlová, and Karel Hornák. 2011. “Alga-Derived Substrates Select for Distinct Betaproteobacterial Lineages and Contribute to Niche Separation in Limnohabitans Strains.” *Applied and Environmental Microbiology* 77 (20): 7307–15. https://doi.org/10.1128/AEM.05107-11.

Šimek, Karel, Jirí Nedoma, Jakob Pernthaler, Thomas Posch, and John R. Dolan. 2002. “Altering the Balance between Bacterial Production and Protistan Bacterivory Triggers Shifts in Freshwater Bacterial Community Composition.” *Antonie van Leeuwenhoek* 81 (1/4): 453–63. https://doi.org/10.1023/A:1020557221798.

Solomon, Christopher T., Denise A. Bruesewitz, David C. Richardson, Kevin C. Rose, Matthew C. Van de Bogert, Paul C. Hanson, Timothy K. Kratz, et al. 2013. “Ecosystem Respiration: Drivers of Daily Variability and Background Respiration in Lakes around the Globe.” *Limnology and Oceanography* 58 (3): 849–66. https://doi.org/10.4319/lo.2013.58.3.0849.

Sommaruga, R, I Obernosterer, G J Herndl, and R Psenner. 1997. “Inhibitory Effect of Solar Radiation on Thymidine and Leucine Incorporation by Freshwater and Marine Bacterioplankton.” *Applied and Environmental Microbiology* 63 (11): 4178–84. http://www.ncbi.nlm.nih.gov/pubmed/16535724.

Tsementzi, Despina, Rachel Poretsky, Luis M. Rodriguez-R, Chengwei Luo, and Konstantinos T. Konstantinidis. 2014. “Evaluation of Metatranscriptomic Protocols and Application to the Study of Freshwater Microbial Communities.” *Environmental Microbiology Reports* 6 (6): 640–55. https://doi.org/10.1111/1758-2229.12180.

Vila-Costa, Maria, Shalabh Sharma, Mary Ann Moran, and Emilio O. Casamayor. 2013. “Diel Gene Expression Profiles of a Phosphorus Limited Mountain Lake Using Metatranscriptomics.” *Environmental Microbiology* 15 (4): 1190–1203. https://doi.org/10.1111/1462-2920.12033.

Supp Fig SX showing Top 10 most expressed genes from each study site: for SP, break the y-axis so the genes way below PS P680 gene are not so compressed. You can use an arrow to call attention to the fact that the PS P680 gene is so much higher.

For these graphs we need to make sure the individual point/circle labels are legible. Maybe have a list of the genes on the right, and put numbers in the circles? Better to have the gene names right there on the figure (ie not a separate table) but readers could quickly link the circle with the gene name, but we don't have to do gymnastics to get the names all visible.

Abundance versus expression plots. These are some of the most informative plots. Since you are condensing Figs 2,3,4 into one figure, can we move them to main text?

Supp figure on **Assessing the variability of metatranscriptomic read counts.** I sort of understand eigenvectors in the context of multivariate ordinations, but I don't understand where these are coming from here?