# Time-series metatranscriptomes reveal conserved patterns and ecological interactions between phototrophic and heterotrophic microbes in diverse freshwater systems

Alexandra M. Linz1\*, Frank O. Aylward2, Stefan Bertilsson3, Katherine D. McMahon4,5

1Great Lakes Bioenergy Research Center, 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, 5Department of Bacteriology, University of Wisconsin–Madison

\*Corresponding author

## Abstract

Microbial communities form the base of freshwater ecosystems, yet the interactions within these diverse communities are poorly understood. Based on evidence showing that primary production and respiration follow diurnal trends in lakes, we hypothesized that gene expression in freshwater microbes would have similar diel cycles, regardless of variation in lake characteristics. We used three two-day time series of metatranscriptomes to test this hypothesis in a eutrophic lake, an oligotrophic lake, and a humic lake. Using this dataset, we identified both differential expression in day versus night and diel cycles in all three lakes. Specifically, genes related to photosynthesis were more expressed during daylight, and genes related to sugar transport were more expressed at night, suggesting that primary production and respiration are not only performed by different community members, but also performed at different times. These results indicate sophisticated organization within freshwater microbial communities that is generalizable across lake types.

**Scientific Significance Statement**

The importance and diversity of microbial communities in aquatic ecosystems has become increasingly apparent as next-generation sequencing techniques provide extensive data on unculturable microbes. Still, one of the grand challenges in aquatic microbiology is linking taxonomic groups to their functions and understanding how they function in a community. In this study, we generated one of the largest metatranscriptomic datasets to date and used it to infer function and interactions between microbes based on trends in gene expression. The results of our work shed light on how organization within microbial communities leads to trends that are detectable at the ecosystem level.

**Data Availability**

Datasets used in this study are available on the Open Science Framework (DOI pending). All code is available at <https://github.com/McMahonLab/geodes>. Raw sequence files are available through the JGI Genome Portal; IDs linked to each metagenome, metatranscriptome, and single amplified genome are provided in Table S1.

**Keywords**

Microbial communities, metatranscriptomics, diel cycles

## Introduction

Many of the core ecosystem functions in freshwater lakes are driven by microbial communities. While the impact of each cell is miniscule, their collective actions form a dynamic, interconnected community whose emergent functions are visible on the ecosystem-level (Goldford et al., 2018; Sunagawa et al., 2015). Previous research in a wide range of freshwater ecosystems indicates that diurnal cycles drive photosynthesis, respiration, and dissolved organic matter (DOM) concentrations (Bertilsson & Jones, 2003; Kaplan & Bott, 1989; Solomon et al., 2013). This implies metabolic interactions between the phototrophic (photosynthetic, also known as phytoplankton) and heterotrophic (non-photosynthetic, also known as bacterioplankton) microbial communities. We hypothesized that these diel trends would be reflected in gene expression. Therefore, we investigated the mechanisms of specific community interactions using the timing of gene expression across the community. To this end, we produced three two-day time series of metatranscriptomes from three lakes with contrasting biogeochemistry. We hypothesized that diel trends in gene expression occur in both phototrophs and heterotrophs due to the direct impacts of sunlight (such as photosynthesis and reactive oxygen species) and its indirect effects (such as metabolite exchange), regardless of the features that distinguish individual lakes.

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, Sharma, Moran, & Casamayor, 2013). Another study in marine systems also observed enhanced 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 phototrophic gene expression followed by a cascade of heterotrophic gene expression have been observed in marine systems (Ottesen et al., 2014). Furthermore, patterns in transcriptional networks of gene expression were consistent in two different regions of the Pacific Ocean, potentially indicating that linkages between phototrophs and heterotrophs are a generalizable feature of marine microbial communities (Aylward et al., 2015). These studies suggest that diel transcriptional trends may be a universal characteristic of both phototrophs and heterotrophs in aquatic microbial communities.

Other methods also suggest strong connections between phototrophs and heterotrophs in aquatic ecosystems. Co-cultures of phototrophic algae and heterotrophic bacteria are often stable over time, indicating mutualistic interactions, although competition or predation have also been described in the laboratory (Cole, 1982; Pernthaler et al., 2001; Posch et al., 1999). In both marine and freshwater systems, the compositions of phototrophic and heterotrophic communities are inextricably linked (Paver et al., 2013; Paver, Youngblut, Whitaker, & Kent, 2015; Verity et al., 1999). Perturbations in one portion of the community have been shown to quickly ripple through the rest (Kent et al., 2006; Šimek, Nedoma, Pernthaler, Posch, & Dolan, 2002; Sjöstedt et al., 2012). One potential mechanism that can explain these trends is DOM release by phototrophs. This DOM is to a large extent composed of low molecular weight compounds, such as sugars, amino acids, carboxylic acids, and alditols (Hellebust, 1965; Maršálek & Rojíčková, 1996). Up to 80% photosynthetic carbon is released extracellularly in marine systems, while up to 99% release has been reported in freshwaters (Bertilsson & Jones, 2003). Although factors causing 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 DOM (Simek, Kasalický, Zapomĕlová, & Hornák, 2011).

Exposure to solar radiation may be a major factor driving gene expression in aquatic ecosystems. Beyond the canonical oxygenic photosynthesis, the presence of opsins, extensively documented in both freshwater and marine heterotrophs, may also lead to cycles of diel gene expression in microbes (Atamna-Ismaeel et al., 2008; Pinhassi, DeLong, Béjà, González, & Pedrós-Alió, 2016). 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). Photodegradation of complex DOM into more labile forms is another potential mechanism that could drive diel trends in heterotrophs (Bertilsson & Tranvik, 2000; Jorgenson, Tranvik, Edling, Graneli, & Lindell, 1998). Sunlight also causes oxidative stress, and heterotrophs may time their metabolisms to avoid this stress (Sommaruga, Obernosterer, Herndl, & Psenner, 1997).

To identify generalizable diel interactions in freshwater microbial communities, we sequenced metatranscriptomes from the epilimnia of three freshwater lakes representing oligotrophic, eutrophic, and dystrophic (humic) lake types. These metatranscriptomes form a two-day time series for each lake, with samples collected every four hours. We additionally sequenced metagenomes and single-cell amplifed genomes (SAGs) from each lake to generate highly specific references, allowing us to obtain higher quality annotations and classifications than possible through read-based annotations. We observed diel trends in both phototrophs and heterotrophs and were able to propose biotic and abiotic mechanisms for these trends based on annotations of expressed genes. Although different taxa and genes were expressed in the three lakes studied, we identified diel trends in all sites, particularly in genes related to photosynthesis and sugar transport.

## Methods

### Study design and in situ measurements

Three lakes in Wisconsin, USA, were chosen for this study based on their different 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. Therefore, 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 microbial studies have been performed in all three lakes, providing reference genomes specific to each site (Bendall et al., 2016; Ghylin et al., 2014; Linz et al., 2018).

Hereafter, we provide brief summaries of our methods; full protocols are available in Document S1. The epilimnion (top thermal layer) of each lake was sampled twelve times at four-hour intervals in July 2016. We used an instrumented sonde (Hydrolab DS5X, OTT Hydromet, Kempten, Germany) equipped with sensors for temperature, dissolved oxygen concentrations, pH, conductivity, and turbidity to collect measurements of the epilimnion. Photosynthetically active radiation (PAR) was also measured at this time using a PAR meter (Li-Cor, Lincoln, NE, USA). Secchi depth was measured once per lake during the time series.

At each time point, 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 in each lake. To collect RNA, water from the integrated epilimnion sample was pumped through 0.22-m polyethylene filters (Pall, Port Washington, NY, USA). Filters were flash frozen in liquid nitrogen in the field and stored at -80oC until extraction and sequencing. Additional samples were collected for metagenomic sequencing, single cell sequencing, total and dissolved nitrogen and phosphorus concentrations, chlorophyll concentrations, and bacterial production assays using 14C-leucine (Chin-Leo & Kirchman, 1988).

### RNA extraction

Samples were lysed with EDTA and SDS and incubated at 65oC, then subjected to bead-beating (FastDNA Spin Kit for Soil, MP Biomedicals, Santa Ana, CA, USA) with TRIzol (Thermo-Fisher, Waltham, MA, USA). An internal standard - an *in vitro* transcription of the cloning plasmid pFN18A was added to samples after beadbeating (Satinsky, Gifford, Crump, & Moran, 2013). Phenol:chloroform was used to isolate RNA from the lysate. Purified 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.

### Additional lab-based measurements

Chlorophyll was extracted with methanol from frozen filters and subsequently acidified to measure phaeophytin. Total and dissolved nitrogen and phosphorus were measured with a colorimetric autoanalyzer. DNA was extracted using phenol:chloroform and the same lysis method as in the RNA extraction protocol. Four additional DNA samples collected from SL in a similar manner in 2009 used as additional references for this lake.

*Reference genomes*

Single-cell amplified genomes were generated following the Department of Energy Joint Genome Institute’s (JGI) standard protocol (Rinke et al., 2014). Briefly, individual cells were sorted using an Influx flow cytometer (BD Biosciences) and treated with Ready-Lyse lysozyme (Epicentre; 5U/μl final concentration) for 15min at room temperature.  Next, cell lysis and whole-genome amplification was performed with the REPLI-g Single Cell Kit (Qiagen) in 2μl reactions.  Lysis and stop reagents from the REPLI-g kit received UV treatment to remove potential DNA contamination (Woyke et al., 2011). Cells for SAG sequencing were chosen with a preference for SL, 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 (Illumina, San Diego, CA, USA). Sequencing reads were filtered using BBTools (Bushnell et al., 2014) and assembled into SAGs using SPAdes (Bankevich et al., 2012).

Metagenomes were prepared for sequencing using the KAPA-Illumina library creation kit (KAPA Biosystems). Metagenomes were sequenced on the Illumina HiSeq platform utilizing a TruSeq paired-end cluster kit (Illumina, San Diego, CA, USA), producing paired ends of 150bp (2x150). Quality filtering was performed on the resulting reads before assembly. BBDuk adapter trimming was used to remove known Illumina adapters (Bushnell et al., 2014). Reads ends were trimmed where quality values were less than 12. Read pairs containing more than three 'N', or with quality scores (before trimming) averaging less than 3 over the read, or length under 51bp after trimming were discarded. Filtered reads were assembled using MegaHit (Li et al., 2016) with a range of kmers (--k-list 23, 43, 63, 83, 103, 123) and otherwise default settings.

Assembled metagenomic contigs, newly sequenced SAGs, genomes from previous McMahon Lab time series sequencing on these lakes (Garcia et al., 2018; Ghylin et al., 2014; Linz et al., 2018), and freshwater algal genomes from NCBI RefSeq (Pruitt & Maglott, 2001) were used to build a nonredundant, highly specific database for subsequent mapping of metatranscriptomic reads (Table S1). 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, Niu, Gao, Fu, & Li, 2010). The longest gene in each cluster was chosen as the representative sequence and used as the mapping reference.

Individual metagenome assemblies were binned using Metabat (Kang, Froula, Egan, & Wang, 2015) and checked for completeness and contamination using CheckM (Parks, Imelfort, Skennerton, Hugenholtz, & Tyson, 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 using in-house McMahon Lab scripts. Where contigs were too short to classify or had conflicting coding region classifications, the coding region classifications were used instead.

Sequencing data for metagenomic reads, assembled metagenomes, and SAGs are available through the JGI Genome Portal (Table S1).

### Metatranscriptomics

Metatranscriptomic samples were sequenced by the JGI. Ribosomal RNA was depleted using the Illumina Ribo-Zero rRNA Removal Kit, and samples were prepared for sequencing with the Illumina TruSeq Stranded Total RNA HT kit. Samples were sequenced using the Illumina HiSeq platform and a TruSeq paired-end clustering kit for paired-end, 150bp sequencing (2x150). BBDuk adapter trimming was used to remove known Illumina adapters. Reads ends were trimmed where quality values were less than 12. Read pairs containing more than three 'N', or with quality scores (before trimming) averaging less than 3 over the read, or length under 51bp after trimming were discarded. Raw sequencing data for metatranscriptomic reads is available through the JGI Genome Portal (Table S1).

Ribosomal RNA reads were removed using SortMeRNA (Kopylova, Noé, & Touzet, 2012). 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, & Shi, 2014).

Addition of an internal RNA standard allowed for both normalization of expressed reads to transcripts per liter 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 SL, 30 from LM, and 21 from TB. Many samples from day two in the TB time series failed to meet quality control standards.

### Statistics

The statistical software R was used for expression analysis (R Core Team, 2018). To reduce noise in the dataset, the top 20,000 expressed genes in each lake were retained for 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 DESeq2 to test for differential expression (Love, Anders, & Huber, 2016). 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. In addition to normalizing by the internal standard, samples were also normalized using a negative binomial distribution using DESeq2 to control for compositional bias before testing for differential expression (Anders & Huber, 2010). RAIN was used to detect cyclic trends in gene expression (Thaben & Westermark, 2014). Based on PAR, day time points were considered to be 9AM, 1PM, and 5PM, while night time points were considered to be 9PM, 1AM, and 5AM. Results were plotted using the R packages ggplot2 (Wickham, 2009) and cowplot (Wilke, 2017).

## Results

### What genes were expressed?

We first asked which genes were most expressed in each lake across all time points (Figure 1). 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 LM and TB. These genes were most frequently attributed to *Cyanobacteria.*

Because of high expression of genes related to phototrophy in all sites, we also ran this analysis excluding genes associated with phototrophy and unannotated genes (Figure 1). 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 non-photosynthetic genes in LM belonged to *Actinobacteria* acI, including a sugar transporter. In TB, *Verrucomicrobia* and *Armatimonadetes* (formerly candidate phylum OP10) contributed some of the most expressed genes, while in SL, a chaperonin expressed by *Deltaproteobacteria* was among most highly expressed genes. Cytochrome subunits were highly expressed in all lakes, ranking in the top 10 in TB and SL, and in the top 25 in LM.

*Which taxa were expressing genes?*

We next aggregated expressed genes by taxonomic classifications to compare the most expressed taxa to the most abundant taxa based on metagenomic data (Figure 2). We used the same reference database to map metatranscriptomes and metagenomes, making such comparisons possible. No positive trend between gene expression and taxonomic abundance was observed. At the phylum level, *Cyanobacteria* were highly expressed in all three lakes, while viruses were also present, but expressed at low levels for all sites. At the clade level, members of *Actinobacteria* acI were both expressed and abundant, as was *Bacteroidetes* bacI-A in LM and SL. The clade acI-B was particularly abundant and also expressed genes in TB, consistent with previous research identifying acI-B2 as an acidic lake specialist (Newton, Jones, Helmus, & McMahon, 2007).

### Assessing variability in freshwater metatranscriptomes

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 future study designs (Tsementzi et al., 2014). We used the coefficient of variation (CoV), i.e. the ratio of standard deviation to average expression (%), to compare the amount of variability within replicate samples to the variation observed across different time points (Figure S1). Higher CoVs were observed across samples than within the replicates. Still, the upper limit for CoV within replicates approached 200%. This result highlights the importance of replication in metatranscriptomic studies.

### Trends in environmental variables

We examined a suite of potentially relevant environmental variables to compare trends in these to the dynamic shifts observed in gene expression, expecting that several of these trends would be diel. PAR data was used to classify time points as night or day (Figure S2). Parameters that reflect the boundaries between layers within the water column, such as dissolved oxygen, temperature, pH, and conductivity, were strongly diel in LM, but less so in SL and TB (Figure S3). Chlorophyll concentrations, often used as an indicator of primary production, were diel in TB, 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 (Figure S4). No diel trends were observed in total and dissolved nitrogen or phosphorus concentrations.

### Gene expression in day vs. night

To test differential expression in day vs. night, we aggregated time points by day (9AM, 1PM, and 5PM) or night (9PM, 1AM, and 5AM). We identified many 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 pre-screened for differential expression in day vs. night.

Genes related to photosynthesis were significantly more expressed in day vs. night and were likely to be cyclic in all lakes (Figure 3). In LM (Table 2), this expression was largely derived from *Cyanobacteria,* while photosynthesis-related gene expression in TB (Table 3) and SL (Table 4) 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 TB, where it was 7-fold higher during the day; the associated gene clusters were largely taxonomically unclassified due to poor reference matching.

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 (R/S/M) transporters were significantly more expressed at night in LM (Figure 4). General sugar transporters were expressed by *Actinobacteria* (acI-B1 and acTH1-A1)*, Cyanobacteria* (*Synechococcaceae* and *Microcystis*)*,* and *Bacteroidetes,* with a lower proportion of reads derived from *Cyanobacteria* at night compared to day. Ribose transporters and R/S/M transporters were mostly classified as *Actinobacteria* and *Bacteroidetes,* with little difference in profiles between day and night. In TB, genes annotated as transporters for general sugars, ribose, and xylose were significantly more expressed at night. *Actinobacteria* (acI-B)contributed the majority of expressed reads for all three types of sugar transporters, while *Alphaproteobacteria* wasidentified in xylose and general sugar transport, and *Cyanobacteria* and *Armatimonadetes* contributed to general sugar transport. The only significant differentially expressed sugar transport group in SL was R/S/M transport in *Actinobacteria*, although several other types were near the significance threshold.

Reactive oxygen species (ROS) defense is a critical function for microbes exposed to UV radiation. 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 LM and SL. Phyla expressing ROS defense-related genes in LM included *Cyanobacteria, Deltaproteobacteria, Planctomycetes, Verrucomicrobia, Betaproteobacteria, and Bacteroidetes.* In TB, ROS defense-related reads were assigned to *Actinobacteria, Alphaproteobacteria, Armatimonadetes, Bacteroidetes, Betaproteobacteria, Eukaryota, Gammaproteobacteria,* and *Verrucomicrobia,* with roughly a 3rd of reads mapping to unclassified gene clusters. ROS defense gene expression in SL was contributed by *Actinobacterica, Bacteroidetes, Betaproteobacteria, Cyanobacteria, Deltaproteobacteria, Eukaryota, Gammaproteobacteria,* and *Planctomycetes.* Interestingly, *Cyanobacteria* expression of ROS defense genes was comparatively low compared to other groups in both TB and SL.

Several functional gene categories differed in significance between lakes. Genes related to opsins, classified as *Actinobacteria* (acI-B1)and *Bacteroidetes,* were significantly more expressed in day in only LM, as were genes related to proteases (*Cyanobacteria, Betaproteobacteria*, *Gammaproteobacteria, Bacteroidetes,* and *Actinobacteria,* with a higher proportion of *Cyanobacteria* in daytime). Genes associated with xylose transport and RuBisCO were significantly differentially expressed only in TB. Additionally, amino acid transport genes were more expressed at night in TB, classified *Actinobacteria, Alphaproteobacteria, Armatimonadetes,* and *Betaproteobacteria*. Fewer functional groups were differentially expressed in SL compared to the other two lakes, with no groups found to be significant only in SL.

## Discussion

In this study, we sought to identify generalizable interactions in freshwater microbial communities through diel gene expression. Using metatranscriptomic time series, we were able to detect genes that were differentially expressed in day vs. night and identify those with cyclic trends. The functional annotations of those genes allowed us to hypothesize metabolic exchanges in freshwater microbial communities. We found similar trends in the timing and types of genes expressed across all lakes studied, regardless of trophic status.

The balance of primary production and respiration is of interest to those seeking to create carbon budgets for freshwater lakes. Previous research 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 highly expressed in day and often cyclic. Photosynthesis and carbon fixation are often considered to be coupled in the process of primary production; however, we only saw differential expression of RuBisCO genes in TB. Still, expression does not necessarily correlate to direct protein abundance or function (Moran et al., 2013), indicating only that transcriptional regulation of RuBisCO is not diel in LM or SL.

Respiration is a broad category that encompasses the degradation of many carbon substrates. To identify the 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). In all three lakes studied, we found significantly higher expression of genes related to sugar transport at night compared to day. Phototrophs are known to exude sugars (Maršálek & Rojíčková, 1996), suggesting that sugars may be exchanged between phototrophs and heterotrophs.

We studied lakes representing eutrophic (LM), humic (TB), and oligotrophic (SL) trophic statuses. There were key differences in gene expression between these lakes, suggesting that microbial communities function differently under varying nutrient concentrations. For example, genes related to amino acid transport were only differentially expressed in day vs. night in TB, while genes related to opsins were only differentially expressed in LM. Although sugar expression was commonly observed, the types of sugars that were differentially expressed in day vs. night differed by lake. In LM, these genes encode ribose, general sugar, and raffinose/stachyose/melibiose transport, while for TB, they encoded ribose, general sugar, and xylose transport. Among the sugar transporters, only genes encoding raffinose/stachyose/melibiose transport were significantly differentially expressed in SP. Still, the similarities between these three lakes in sugar transport and photosynthesis diel expression indicate a potentially generalizable trend in freshwater.

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 such diel release of sugars has not been observed, day/night partitioning of photosynthesis and sugar metabolism are known to occur in phototrophs (Masuda et al., 2018; Welkie et al., 2018). This diel trend in single cells may extend to community-level interactions. The other hypothesis is that oxidative stress prevents heterotrophs from consuming sugar during the day, even if it is available. We observed significant differential expression in genes related to ROS defense, with higher expression during the day, in all three lakes. Although solar radiation is known to produce low molecular organic acids from high molecular weight organic matter, we did not observe differential expression of genes encoding transporters for typical photodegradation products, such as glycolate or carboxylic acids.

There is ample evidence in marine microbial communities suggesting that carbon released by phototrophs influences heterotrophic community composition to improve phototroph fitness. 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). In marine microbial communities, heterotrophic bacterioplankton are highly dependent upon *Prochlorococcus* exudates and likely perform a critical community function in return, such as the detoxification of hydrogen peroxide or free radicals (Morris, Johnson, Szul, Keller, & Zinser, 2011). *Prochlorococcus* has lost its genes for reactive oxygen species defense and depends on the associated heterotrophic bacteria to supply this function (Ma, Calfee, Morris, Johnson, & Zinser, 2018; Morris, Johnson, Wilhelm, & Zinser, 2016). *Prochlorococcus* likely exudates carbon to maintain redox balance, as it generates more reducing power via photosynthesis than it can allocate to anabolic processes (Bertilsson, Berglund, Pullin, & Chisholm, 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, & Chisholm, 2017)*.*

It is therefore reasonable to hypothesize that freshwater photoautotrophs may be releasing carbohydrates to shape the heterotrophic community, which in turn may benefit phototrophs. Most likely, heterotrophs perform functions that benefit the community, such as ROS defense and vitamin production. The origin of metabolic exchanges that lead to co-dependencies has been postulated to be an important driver of evolution in aquatic communities, as in the “Black Queen Hypothesis” (Morris, Lenski, & Zinser, 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 (Braakman et al., 2017). 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) as drivers of community-level diel trends in freshwater microbiomes. Whether all of these microbes are responding to the same day-night stimulus or whether community interactions confer these diel trends remains to be determined. Given the consistent patterns across metatranscriptomes from biogeochemically disparate lakes, our results underscore the prevalence of conserved microbial interactions that underpin a broad diversity of freshwater environments.

**Funding**

This research was supported by the U.S. Department of Energy Joint Genome Institute through the Community Sequencing Program (CSP 1977). The work conducted by the U.S. Department of Energy Joint Genome Institute, a DOE Office of Science User Facility, is supported by the Office of Science of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231. Katherine D. McMahon received funding from the United States National Science Foundation Microbial Observatories program (MCB-0702395), the Long-Term Ecological Research Program (NTL–LTER DEB-1440297), and an INSPIRE award (DEB-1344254). Alexandra M. Linz was supported by a pre-doctoral fellowship provided by the University of Wisconsin–Madison Department of Bacteriology and by the National Science Foundation Graduate Research Fellowship Program under grant no. DGE-1256259 during this research.

**Acknowledgements**

This work would not have been possible without the large team of volunteers who helped collect and process samples. These people included Amelia Flannery, Benjamin Peterson, Carolyn Voter, Carolyn Pugh, Christine Tam, Christopher Lawson, Dominick Ciruzzi, Francisco Moya, Grace Schrader, Joshua Breider, Katy Forest, Kylie Huang, Madeleine Hughan, Madeleine Magee, Margaret Sobolewski, Mark Gahler, Mykala Sobieck, Pamela Camejo, Robin Rohwer, Sarah Stevens, and Shaomei He. We also thank Sarah Stevens for her contribution of in-house McMahon Lab scripts and her advice throughout the data analysis process. Finally, we thank the Long-Term Ecological Research Program, the Center for Limnology, and UW Trout Lake Station for providing logistical support and equipment during the field sampling.

## References

Anders, S., & Huber, W. (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, N., Sabehi, G., Sharon, I., Witzel, K.-P., Labrenz, M., Jürgens, K., … Beja, O. (2008). Widespread distribution of proteorhodopsins in freshwater and brackish ecosystems. *The ISME Journal*, *2*(6), 656–662. https://doi.org/10.1038/ismej.2008.27

Aylward, F. O., Eppley, J. M., Smith, J. M., Chavez, F. P., Scholin, C. A., & DeLong, E. F. (2015). Microbial community transcriptional networks are conserved in three domains at ocean basin scales. *Proceedings of the National Academy of Sciences*, *112*(17), 5443–5448. https://doi.org/10.1073/PNAS.1502883112

Bankevich, A., Nurk, S., Antipov, D., Gurevich, A. A., Dvorkin, M., Kulikov, A. S., … Pevzner, P. A. (2012). SPAdes: A New Genome Assembly Algorithm and Its Applications to Single-Cell Sequencing. *Journal of Computational Biology*, *19*(5), 455–477. https://doi.org/10.1089/cmb.2012.0021

Bendall, M. L., Stevens, S. L., Chan, L.-K., Malfatti, S., Schwientek, P., Tremblay, J., … Malmstrom, R. R. (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, S., Berglund, O., Pullin, M. J., & Chisholm, S. W. (2005). Release of dissolved organic matter by Prochlorococcus. *Vie et Milieu*, *55*(3–4), 225–232. Retrieved from https://www.researchgate.net/publication/252613206

Bertilsson, S., & Jones, J. B. (2003). Supply of Dissolved Organic Matter to Aquatic Ecosystems: Autochthonous Sources. *Aquatic Ecosystems*, 3–24. https://doi.org/10.1016/B978-012256371-3/50002-0

Bertilsson, S., & Tranvik, L. J. (2000). Photochemical transformation of dissolved organic matter in lakes. *Limnology and Oceanography*, *45*(4), 753–762. https://doi.org/10.4319/lo.2000.45.4.0753

Braakman, R., Follows, M. J., & Chisholm, S. W. (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–E3100. https://doi.org/10.1073/pnas.1619573114

Bushnell, B., Egan, R., Copeland, A., Foster, B., Clum, A., Sun, H., … Scholz, M. (2014). BBMap: A Fast, Accurate, Splice-Aware Aligner. https://doi.org/10.1186/1471-2105-13-238

Chin-Leo, G., & Kirchman, D. L. (1988). Estimating bacterial production in marine waters from the simultaneous incorporation of thymidine and leucine. *Applied and Environmental Microbiology*, *54*(8), 1934–1939. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/16347706

Cole, J. J. (1982). *INTERACTIONS BETWEEN BACTERIA AND ALGAE IN AQUATIC ECOSYSTEMS*. *Ann Rev. Ecol. Syst* (Vol. 13). Retrieved from www.annualreviews.org

Garcia, S. L., Stevens, S. L. R., Crary, B., Martinez-Garcia, M., Stepanauskas, R., Woyke, T., … McMahon, K. D. (2018). Contrasting patterns of genome-level diversity across distinct co-occurring bacterial populations. *The ISME Journal*, *12*(3), 742–755. https://doi.org/10.1038/s41396-017-0001-0

Ghylin, T. W., Garcia, S. L., Moya, F., Oyserman, B. O., Schwientek, P., Forest, K. T., … McMahon, K. D. (2014). Comparative single-cell genomics reveals potential ecological niches for the freshwater acI Actinobacteria lineage. *The ISME Journal*, *8*(12), 2503–2516. https://doi.org/10.1038/ismej.2014.135

Goldford, J. E., Lu, N., Bajić, D., Estrela, S., Tikhonov, M., Sanchez-Gorostiaga, A., … Sanchez, A. (2018). Emergent simplicity in microbial community assembly. *Science (New York, N.Y.)*, *361*(6401), 469–474. https://doi.org/10.1126/science.aat1168

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, Y., Niu, B., Gao, Y., Fu, L., & Li, W. (2010). CD-HIT Suite: a web server for clustering and comparing biological sequences. *Bioinformatics*, *26*(5), 680–682. https://doi.org/10.1093/bioinformatics/btq003

Jorgenson, N. O., Tranvik, L. J., Edling, H., Graneli, W., & Lindell, M. (1998). Effects of sunlight on occurrence and bacterial turnover of specific carbon and nitrogen compounds in lake water. *FEMS Microbiology Ecology*, *25*, 217–227.

Kang, D. D., Froula, J., Egan, R., & Wang, Z. (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, L. A., & Bott, T. L. (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), 718–733. https://doi.org/10.4319/lo.1989.34.4.0718

Kent, A. D., Jones, S. E., Lauster, G. H., Graham, J. M., Newton, R. J., & McMahon, K. D. (2006). Experimental manipulations of microbial food web interactions in a humic lake: shifting biological drivers of bacterial community structure. *Environmental Microbiology*, *8*(8), 1448–1459. https://doi.org/10.1111/j.1462-2920.2006.01039.x

Kopylova, E., Noé, L., & Touzet, H. (2012). SortMeRNA: fast and accurate filtering of ribosomal RNAs in metatranscriptomic data. *Bioinformatics*, *28*(24), 3211–3217. https://doi.org/10.1093/bioinformatics/bts611

Li, D., Luo, R., Liu, C.-M., Leung, C.-M., Ting, H.-F., Sadakane, K., … Lam, T.-W. (2016). MEGAHIT v1.0: A fast and scalable metagenome assembler driven by advanced methodologies and community practices. *Methods*, *102*, 3–11. https://doi.org/10.1016/J.YMETH.2016.02.020

Liao, Y., Smyth, G. K., & Shi, W. (2014). featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics*, *30*(7), 923–930. https://doi.org/10.1093/bioinformatics/btt656

Linz, A. M., He, S., Stevens, S. L. R., Anantharaman, K., Rohwer, R. R., Malmstrom, R. R., … McMahon, K. D. (2018). Freshwater carbon and nutrient cycles revealed through reconstructed population genomes. *PeerJ*, *6*, e6075. https://doi.org/10.7717/peerj.6075

Love, M. I., Anders, S., & Huber, W. (2016). Differential analysis of count data-the DESeq2 package. https://doi.org/10.1186/s13059-014-0550-8

Ma, L., Calfee, B. C., Morris, J. J., Johnson, Z. I., & Zinser, E. R. (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), 473–484. https://doi.org/10.1038/ismej.2017.182

Maresca, J. A., Keffer, J. L., Hempel, P., Polson, S. W., Shevchenko, O., Bhavsar, J., … Hahn, M. W. (2019). Light modulates the physiology of non-phototrophic Actinobacteria. *Journal of Bacteriology*, JB.00740-18. https://doi.org/10.1128/JB.00740-18

Markowitz, V. M., Chen, I. M. A., Palaniappan, K., Chu, K., Szeto, E., Grechkin, Y., … Kyrpides, N. C. (2012). IMG: The integrated microbial genomes database and comparative analysis system. *Nucleic Acids Research*, *40*(D1), 115–122. https://doi.org/10.1093/nar/gkr1044

Maršálek, B., & Rojíčková, R. (1996). Stress Factors Enhancing Production of Algal Exudates: a Potential Self-Protective Mechanism? *Zeitschrift Für Naturforschung C*, *51*(9–10), 646–650. https://doi.org/10.1515/znc-1996-9-1008

Masuda, T., Bernát, G., Bečková, M., Kotabová, E., Lawrenz, E., Lukeš, M., … Prášil, O. (2018). Diel regulation of photosynthetic activity in the oceanic unicellular diazotrophic cyanobacterium *Crocosphaera watsonii* WH8501. *Environmental Microbiology*, *20*(2), 546–560. https://doi.org/10.1111/1462-2920.13963

Moran, M. A., Satinsky, B., Gifford, S. M., Luo, H., Rivers, A., Chan, L.-K., … Hopkinson, B. M. (2013). Sizing up metatranscriptomics. *The ISME Journal*, *7*(2), 237–243. https://doi.org/10.1038/ismej.2012.94

Morris, J. J., Johnson, Z. I., Szul, M. J., Keller, M., & Zinser, E. R. (2011). Dependence of the Cyanobacterium Prochlorococcus on Hydrogen Peroxide Scavenging Microbes for Growth at the Ocean’s Surface. *PLoS ONE*, *6*(2), e16805. https://doi.org/10.1371/journal.pone.0016805

Morris, J. J., Johnson, Z. I., Wilhelm, S. W., & Zinser, E. R. (2016). Diel regulation of hydrogen peroxide defenses by open ocean microbial communities. *Journal of Plankton Research*, *38*(4), 1103–1114. https://doi.org/10.1093/plankt/fbw016

Morris, J. J., Lenski, R. E., & Zinser, E. R. (2012). The Black Queen Hypothesis: evolution of dependencies through adaptive gene loss. *MBio*, *3*(2), e00036-12. https://doi.org/10.1128/mBio.00036-12

Nelson, C. E., Goldberg, S. J., Wegley Kelly, L., Haas, A. F., Smith, J. E., Rohwer, F., & Carlson, C. A. (2013). Coral and macroalgal exudates vary in neutral sugar composition and differentially enrich reef bacterioplankton lineages. *The ISME Journal*, *7*(5), 962–979. https://doi.org/10.1038/ismej.2012.161

Newton, R. J., Jones, S. E., Helmus, M. R., & McMahon, K. D. (2007). Phylogenetic ecology of the freshwater Actinobacteria acI lineage. *Applied and Environmental Microbiology*, *73*(22), 7169–7176. https://doi.org/10.1128/AEM.00794-07

Ottesen, E. A., Young, C. R., Eppley, J. M., Ryan, J. P., Chavez, F. P., Scholin, C. A., & DeLong, E. F. (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, E. A., Young, C. R., Gifford, S. M., Eppley, J. M., Marin, R., Schuster, S. C., … DeLong, E. F. (2014). Ocean microbes. Multispecies diel transcriptional oscillations in open ocean heterotrophic bacterial assemblages. *Science (New York, N.Y.)*, *345*(6193), 207–212. https://doi.org/10.1126/science.1252476

Parks, D. H., Imelfort, M., Skennerton, C. T., Hugenholtz, P., & Tyson, G. W. (2015). CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Research*, *25*(7), 1043–1055. https://doi.org/10.1101/gr.186072.114

Paver, S. F., Hayek, K. R., Gano, K. A., Fagen, J. R., Brown, C. T., Davis-Richardson, A. G., … Kent, A. D. (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, S. F., Youngblut, N. D., Whitaker, R. J., & Kent, A. D. (2015). Phytoplankton succession affects the composition of *P* *olynucleobacter* subtypes in humic lakes. *Environmental Microbiology*, *17*(3), 816–828. https://doi.org/10.1111/1462-2920.12529

Pernthaler, J., Posch, T., Simek, K., Vrba, J., Pernthaler, A., Glöckner, F. O., … Amann, R. (2001). Predator-specific enrichment of actinobacteria from a cosmopolitan freshwater clade in mixed continuous culture. *Applied and Environmental Microbiology*, *67*(5), 2145–2155. https://doi.org/10.1128/AEM.67.5.2145-2155.2001

Pinhassi, J., DeLong, E. F., Béjà, O., González, J. M., & Pedrós-Alió, C. (2016). Marine Bacterial and Archaeal Ion-Pumping Rhodopsins: Genetic Diversity, Physiology, and Ecology. *Microbiol. Mol. Biol. Rev.*, *80*(4), 929–954. https://doi.org/10.1128/MMBR.00003-16

Poretsky, R. S., Hewson, I., Sun, S., Allen, A. E., Zehr, J. P., & Moran, M. A. (2009). Comparative day/night metatranscriptomic analysis of microbial communities in the North Pacific subtropical gyre. *Environmental Microbiology*, *11*(6), 1358–1375. https://doi.org/10.1111/j.1462-2920.2008.01863.x

Posch, T., Simek, K., Vrba, J., Pernthaler, J., Nedoma, J., Sattler, B., … Psenner, R. (1999). Predator-induced changes of bacterial size-structure and productivity studied on an experimental microbial community. *Aquatic Microbial Ecology*, *18*(3), 235–246. https://doi.org/10.3354/ame018235

Pruitt, K. D., & Maglott, D. R. (2001). RefSeq and LocusLink: NCBI gene-centered resources. *Nucleic Acids Research*, *29*(1), 137–140. https://doi.org/10.1093/nar/29.1.137

Rinke, C., Lee, J., Nath, N., Goudeau, D., Thompson, B., Poulton, N., … Woyke, T. (2014). Obtaining genomes from uncultivated environmental microorganisms using FACS–based single-cell genomics. *Nature Protocols*, *9*(5), 1038–1048. https://doi.org/10.1038/nprot.2014.067

Satinsky, B. M., Gifford, S. M., Crump, B. C., & Moran, M. A. (2013). Use of Internal Standards for Quantitative Metatranscriptome and Metagenome Analysis. *Methods in Enzymology*, *531*, 237–250. https://doi.org/10.1016/B978-0-12-407863-5.00012-5

Simek, K., Kasalický, V., Zapomĕlová, E., & Hornák, K. (2011). Alga-derived substrates select for distinct Betaproteobacterial lineages and contribute to niche separation in Limnohabitans strains. *Applied and Environmental Microbiology*, *77*(20), 7307–7315. https://doi.org/10.1128/AEM.05107-11

Šimek, K., Nedoma, J., Pernthaler, J., Posch, T., & Dolan, J. R. (2002). Altering the balance between bacterial production and protistan bacterivory triggers shifts in freshwater bacterial community composition. *Antonie van Leeuwenhoek*, *81*(1/4), 453–463. https://doi.org/10.1023/A:1020557221798

Sjöstedt, J., Koch-Schmidt, P., Pontarp, M., Canbäck, B., Tunlid, A., Lundberg, P., … Riemann, L. (2012). Recruitment of members from the rare biosphere of marine bacterioplankton communities after an environmental disturbance. *Applied and Environmental Microbiology*, *78*(5), 1361–1369. https://doi.org/10.1128/AEM.05542-11

Solomon, C. T., Bruesewitz, D. A., Richardson, D. C., Rose, K. C., Van de Bogert, M. C., Hanson, P. C., … Zhu, G. (2013). Ecosystem respiration: Drivers of daily variability and background respiration in lakes around the globe. *Limnology and Oceanography*, *58*(3), 849–866. https://doi.org/10.4319/lo.2013.58.3.0849

Sommaruga, R., Obernosterer, I., Herndl, G. J., & Psenner, R. (1997). Inhibitory effect of solar radiation on thymidine and leucine incorporation by freshwater and marine bacterioplankton. *Applied and Environmental Microbiology*, *63*(11), 4178–4184. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/16535724

Sunagawa, S., Coelho, L. P., Chaffron, S., Kultima, J. R., Labadie, K., Salazar, G., … Bork, P. (2015). Structure and function of the global ocean microbiome. *Science (New York, N.Y.)*, *348*(6237), 1261359. https://doi.org/10.1126/science.1261359

Thaben, P. F., & Westermark, P. O. (2014). Detecting rhythms in time series with RAIN. *Journal of Biological Rhythms*, *29*(6), 391–400. https://doi.org/10.1177/0748730414553029

Tsementzi, D., Poretsky, R., Rodriguez-R, L. M., Luo, C., & Konstantinidis, K. T. (2014). Evaluation of metatranscriptomic protocols and application to the study of freshwater microbial communities. *Environmental Microbiology Reports*, *6*(6), 640–655. https://doi.org/10.1111/1758-2229.12180

Verity, P. G., Wassmann, P., Ratkova, T. N., Andreassen, I. J., Nordby, E., & Høisæter, T. (1999). Seasonal patterns in composition and biomass of autotrophic and heterotrophic nano- and microplankton communities on the north Norwegian shelf. *Sarsia*, *84*(3–4), 265–277. https://doi.org/10.1080/00364827.1999.10420431

Vila-Costa, M., Sharma, S., Moran, M. A., & Casamayor, E. O. (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

Welkie, D. G., Rubin, B. E., Diamond, S., Hood, R. D., Savage, D. F., & Golden, S. S. (2018). A Hard Day’s Night: Cyanobacteria in Diel Cycles. *Trends in Microbiology*. https://doi.org/10.1016/J.TIM.2018.11.002

Woyke, T., Sczyrba, A., Lee, J., Rinke, C., Tighe, D., Clingenpeel, S., … Cheng, J.-F. (2011). Decontamination of MDA Reagents for Single Cell Whole Genome Amplification. *PLoS ONE*, *6*(10), e26161. https://doi.org/10.1371/journal.pone.0026161