# 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 of 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 in a dynamic, interconnected community whose functions can impact an entire ecosystem. We used metatranscriptomic sequencing to better understand how microbial community gene expression varies over day versus night in three lakes with contrasting characteristics. Based on previous process rate measurements in a wide range of freshwater ecosystems, we hypothesized that genes related to photoautotrophy (primary production) would be expressed during the day and that genes related to heterotrophic carbon metabolism would be differentially expressed in day vs. night in all three lake types.

The hypothesis that heterotrophic bacteria alter their physiology of the course of a diel cycle may seem counterintuitive, but previous metatranscriptomic work in marine and freshwater systems has identified these patterns and 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 pyrophosphatases [(10)](https://paperpile.com/c/kQRinB/R3gI). 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 [(11)](https://paperpile.com/c/kQRinB/32NM). 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 (cite Otteson) [(12)](https://paperpile.com/c/kQRinB/l3sB). A trend of peak activity in heterotrophic bacterioplankton following that of photoautotrophs was consistent in two different regions of the Pacific Ocean, suggesting strong metabolic links between these two groups in disparatemarine microbial communities. One mechanism that may explain such diel trends in heterotrophic organisms is metabolite exchange between the phototrophic and heterotrophic communities.

Experimental work has indicated that interactions between phototrophs and heterotrophs may be prevalent in aquatic environments. 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 [(2)](https://paperpile.com/c/kQRinB/nZXs). In lake ecosystems, changes in the phototrophic community are known to cause variability also in the heterotrophic community [(3)](https://paperpile.com/c/kQRinB/4WjX), potentially because compounds produced by the phototrophic community (such as glycolate) [(4)](https://paperpile.com/c/kQRinB/O7uE) or decaying biomass can be consumed by heterotrophs. Freshwater phototrophs are also known to release carbohydrates, which can enhance the growth of heterotrophs [(5)](https://paperpile.com/c/kQRinB/ryBs). Up to 25% of fixed carbon can be excreted by algae [(6)](https://paperpile.com/c/kQRinB/jrY2). Some ubiquitous freshwater bacteria, such as *Limnohabitans,* appear to specialize in algal-derived carbon uptake [(7)](https://paperpile.com/c/kQRinB/usa5). Solar driven photodegradation of complex dissolved organic carbon (DOC) [(8)](https://paperpile.com/c/kQRinB/Ucg0) and the presence of rhodopsins [(9)](https://paperpile.com/c/kQRinB/IjYF) could also lead to diel trends in gene expression of heterotrophic bacteria. Given these multiple lines of evidence, we hypothesized that diurnal changes in solar irradiance may be a significant factor driving gene expression also of heterotrophic metabolism.

In this study, we use a high-resolution metatranscriptomic time series to investigate differences in community gene expression in day vs. night. We collected metatranscriptomic samples every four hours for two days and repeated this experiment in three freshwater ecosystems representing oligotrophic, eutrophic, and dystrophic (humic) lake types. This study is one of the largest metatranscriptomic sequencing efforts on freshwater communities to date; therefore, we discuss the variability observed in this dataset to inform the design of future metatranscriptomic studies of freshwater. We detected many significant differences in gene expression in day vs. night in both phototrophic and heterotrophic microbes, and suggest that metabolite exchange can explain the effect of sunlight on heterotrophic metabolisms in lake bacterioplankton.

## 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), eutrophic (Lake Mendota), and humic (Trout Bog). The limnological characteristics of each lake are presented in Table 1.Lake Mendota is located in Madison, WI, USA, while Trout Bog and Sparkling Lake are located in Boulder Junction, WI, USA, approximately four hours 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 [(13, 14)](https://paperpile.com/c/kQRinB/bx1Z+I0sv).

The epilimnion 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 Sparkling Lake and Trout Bog was slightly longer than at Lake Mendota. 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 Trout Bog, 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.

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 were washed with ambient epilimnion water prior to each timepoint. RNA samples were the first samples collected at each timepoint. Water from the integrated epilimnion sample was immediately pumped through 0.22 micron 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, determined for each lake prior to beginning the time series (2-4 minutes), and the volume filtered was recorded. Four replicate filters for RNA extraction were collected from each integrated epilimnion sample. Filters were placed in 2mL plastic cryovials (Phenix Research, Candler, NC, USA) and immediately flash frozen in liquid nitrogen in the field. Filters were stored at -80C after collection, and samples from Trout Bog and Sparkling Lake 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), with samples for dissolved nutrient analysis collected using effluent from the RNA filtration. 1L of unfiltered water for chlorophyll measurement was collected in a black, opaque glass bottle and filtered on shore (approximately 30 minutes after collection). Three replicates of 250 mL from the dark bottle were filtered through a 0.3 micron nitrocellulose Whatman filter (GE Healthcare Life Sciences, Pittsburgh, PA, USA) and flash frozen in liquid nitrogen. 15mL of unfiltered water was collected in 50 mL Falcon tubes (Corning, Corning, NY, USA) for protein synthesis assays using C14-leucine assays. This sample was stored in a thermos of epilimnion water to maintain ambient temperature during transport to designated radioactive lab spaces; this same thermos was used for subsequent incubations during the bacterial production assay. 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 micron filters used for RNA collection. Cells were preserved for single amplified genome sequencing by mixing 2mL 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 -80C until processing.

### RNA extraction

Within 2-3 weeks of collection, RNA was extracted from the filters in a single batch operation. A detailed protocol is available in the supplemental materials; a brief overview is presented here. Filters were exposed to a lysis solution containing EDTA and SDS and incubated at 65C. 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 [(15)](https://paperpile.com/c/kQRinB/EXys). The samples were then centrifuged for five minutes, and the supernatant was transferred to a fresh tube. From this point, the protocol resembles a typical phenol-chloroform extraction. Chloroform was used to separate RNA from other molecules in the TRIzol. After cleaning, the RNA was precipitated in ethanol, pelleted, and resuspended. The RNA was further purified using an RNeasy kit (QIAGEN, Hilden, Germany), which includes DNAse digestion. All samples were quantified using a Qubit fluorimeter (Thermo-Fisher) and stored at -80C 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. Samples were acidified to measure phycocyanin in addition to chlorophyll. Extracted chlorophyll was diluted as needed to remain within range of the spectrometer; samples from Sparkling Lake required no dilution, samples from Lake Mendota required a 1:4 dilution, and samples from Trout Bog required either a 1:2 or a 1:4 dilution.

Bacterial production assays were conducted at each timepoint using 14C-leucine. 1.5 mL of water were added to six microcentrifuge tubes. Two of the six samples were immediately killed using trichloroacetic acid (TCA) as negative controls. All samples received 14C-leucine and were incubated for one hour, after which remaining “live” samples were killed with TCA and stored at -20C. Approximately one month after sample collection, production assay samples were thawed, pelleted, and resuspended in ethanol. Radioactivity was measured using a liquid scintillation counter.

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. RNA samples were sequenced using Illumina HiSeq 2500-1TB (Illumina, San Diego, CA, USA). When samples failed to sequence with sufficient quality, the fourth replicate of those samples was sequenced instead. Metatranscriptomic reads were quality filtered by the JGI. Metatranscriptomic reads were assembled by the JGI using MetaHit [(16)](https://paperpile.com/c/kQRinB/ZxdR). DNA samples for metagenomics were also sequenced on an Illumina HiSeq platform. Metagenomic reads were assembled by the JGI and assembled using MetaHit. 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 [(17)](https://paperpile.com/c/kQRinB/I8ZD).

Cells for single amplified genomes were sorted, identified using 16S 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 Lake, the least well-represented lake in the pre-existing databases. 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 [(18)](https://paperpile.com/c/kQRinB/D8P2). Unscreened SAGs were used as references to retain any unusual DNA sequences in the genome.

### Bioinformatics pipeline

Ribosomal RNA reads, which still comprised approximately 50% of metatranscriptomic reads despite depletion prior to sequencing, were first removed using SortMeRNA [(19)](https://paperpile.com/c/kQRinB/kbRD). Assembled metagenomic contigs from this study, SAGs from this study, SAGs and MAGs from previous McMahon Lab time series sequencing on these lakes [(13, 14)](https://paperpile.com/c/kQRinB/bx1Z+I0sv), and 5 freshwater algal genomes from NCBI RefSeq [(20)](https://paperpile.com/c/kQRinB/fX0s), representing each algal genus, were used to build a nonredundant, highly specific database for subsequent mapping of metatranscriptomic reads. 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 [(21)](https://paperpile.com/c/kQRinB/Pc8N). Metatranscriptomic reads were mapped to this database with a 90% ID cutoff using BBMap. Mapped reads were tabulated using FeatureCounts [(22)](https://paperpile.com/c/kQRinB/Rit1). Metagenome assemblies were binned by lake using Metabat [(23)](https://paperpile.com/c/kQRinB/taMr) and checked for completeness and contamination using CheckM [(24)](https://paperpile.com/c/kQRinB/1ntJ). 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 [(25)](https://paperpile.com/c/kQRinB/fhFA) for each coding region on a contig/bin (Stevens, unpublished).

### Statistics

During the extraction process, we added an internal standard to our RNA samples immediately after cell lysis. This 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 Lake, 30 from Lake Mendota, and 21 from Trout Bog. Many samples from day two in the Trout Bog time series failed to meet to quality control standards, suggesting sampling error for these timepoints. One likely source of the issue is insufficient RNA yield due to gradual slowing of the peristaltic pumps after four days of nearly continuous use. The resulting read counts are in units of transcripts per liter and are semi-quantitative (keeping in mind the inherent limitations and biases of metatranscriptomic sequencing).

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 - for example, any gene annotation containing “nitrogenase,” “nifH|nifD|nifK,” or “nitrogen fixation” was considered representative of the pathway nitrogen fixation. The summed expression of each pathway/process was input into DESeq to test differential expression [(26)](https://paperpile.com/c/kQRinB/VqZR). Despite normalizing by the internal standard, samples were still normalized by size factors to control for compositional bias, as recommended by the authors of DESeq. To account for the large number of pathways tested, an adjusted p-value of less than 0.05 was used as an indicator of significantly different expression. This analysis was run both with lake as a condition and with day or night as a condition within each lake. 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).

## Results

### What genes are 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 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.* A photosystem gene derived from *Chitinophagaceae*, a family of *Bacteroidetes* not known to perform photosynthesis [(27)](https://paperpile.com/c/kQRinB/mk5V),was highly expressed in Trout Bog. As *Bacteroidetes* is proposed to form a superphylum with *Chlorobi*, it is possible that these genes are instead derived from green sulfur bacteria. Interestingly, a hypothetical gene from the bacterial predator *Bdellovibrio* [*(28)*](https://paperpile.com/c/kQRinB/99Cl)and a gene encoding PQQ-dependent dehydrogenase were highly expressed in Lake Mendota.

Because of the dominance of phototrophic taxa and genes at all sites, we also investigated which genes were highly expressed from the heterotrophic component of the microbial community (Table 3). 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, including a sugar transporter. In Trout Bog, *Verrucomicrobia* and *Armatimonadetes* contributed some of the top expressed genes, while in Sparkling Lake, a chaperonin expressed by *Deltaproteobacteria* was the most highly expressed heterotrophic gene.

*What phyla are expressed?*

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 outlier - genes with this classification were almost exclusively derived from a single, low quality MAG.

### Trends in environmental variables

We collected data on many other 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). Concentrations of chlorophyll, 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 all three lakes, although the trends were not diel (supplemental). No trends were observed in nitrogen or phosphorus concentrations. It is unclear based on our metatranscriptomic data which taxa or genes may be driving trends in our measured environmental variables.

### 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, suggesting that much of this variability is biological in origin. Still, the upper limit for CoV within replicates approached 200%. This result highlights the importance of replication in metatranscriptomic studies.

### 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 abundant genes in each lake. While many genes were found to be significantly differentially expressed between day and night in all three lakes, we focused on genes likely to be related to phototroph-heterotroph interactions for the sake of conciseness. However, we do note that previous published metatranscriptomic research has found alkaline phosphatase-encoding genes to be more highly expressed in night vs. day in freshwater. In our metatranscriptomic dataset, Lake Mendota was the only lake where alkaline phosphatase was significantly differentially expressed, and its expression was higher in the day than in the night.

As expected, genes related to photosynthesis were among the most highly expressed genes in all three lakes, and as expected, these were significantly more expressed in day than at night. In Lake Mendota (Figure 3), this expression was largely derived by *Cyanobacteria.*  Genes encoding subunits of the enzyme ribulose-1,5-bisphosphate carboxylase (RuBisCO), the key enzyme in the Calvin-Benson-Bassham (CBB) carbon fixation cycle, were more highly expressed in the day and were also mainly derived from *Cyanobacteria* in Lake Mendota. Expression of photosynthesis genes in Sparkling Lake (Figure 4) were also affiliated with *Cyanobacteria,* with a significant remaining proportion of the reads contributed by *Bacteroidetes.* As previously mentioned, the *Bacteroidetes* family in question, *Chitinophagaceae,* is not known to perform photosynthesis, and these reads may instead be from a misclassified green sulfur bacterium in the proposed phylum *Sphingobacteria* [*(29)*](https://paperpile.com/c/kQRinB/5KOP).

In contrast to Lake Mendota, we observed higher expression of RuBisCO at night than during the day in Sparkling Lake, with all genes taxonomically unclassified. Photosynthesis expression in Trout Bog showed the same trend as for the other two lakes, although a much smaller proportion of photosynthesis expression in Trout Bog were assigned to *Cyanobacteria,* with most of these reads remaining taxonomically unclassified (Figure 5). Genes encoding RuBisCO in Trout Bog were more expressed in day and were taxonomically unclassified.

Another prominent feature of our dataset was that genes related to sugar transport were uniformly and significantly more expressed at night, for all transporter types and across all three lakes. Genes encoding transporters that likely act on chitobiose, fructose, glucose/mannose, lactose/arabinose, raffinose/stachyose/melibiose, rhamnose, ribose, and xylose were differentially expressed in Lake Mendota (Figure 3). Most expressed sugar transport genes in Lake Mendota were classified as Actinobacteria, with a substantial additional proportion classified as Bacteroidetes. The exception to this was genes encoding xylose transport, with most reads derived from *Cyanobacteria.*

Although fewer sugar transporters were found to be significantly differentially expressed in Sparkling Lake than in Lake Mendota, the same pattern of higher expression for these functions at night was consistent (Figure 4). These genes encoded ribose transport in addition to general sugar uptake. Most sugar transporters in Sparkling Lake were classified as coming from *Actinobacteria.* In Trout Bog, expressed sugar transporters likely act on fructose, glucose/mannose, rhamnose, ribose, and xylose, and they were all significantly more expressed at night. Genes encoding fructose and glucose/mannose transport were classified as *Alphaproteobacteria* in Trout Bog, while most other sugar transporters could not be classified (Figure 5). General sugar transporters in Trout Bog were assigned to *Actinobacteria, Alphaproteobacteria,* and *Armatimonadetes,*  although most reads were assigned to unclassified genes.

## Discussion

In this study, we sought to identify how phototrophic and heterotrophic gene expression varies in between day and night in multiple lake types with different biogeochemistry. As expected, genes related to photosynthesis were more highly expressed in the day in all three lakes. Genes encoding the carbon fixation enzyme RuBisCO were also expressed more in the day time in Trout Bog and Lake Mendota, although were more highly expressed at night in Sparkling Lake. This may indicate a disconnect between photosynthesis and carbon fixation in this site, even though these metabolic processes are typically assumed to be tighly coupled in freshwater lakes. Unfortunately, genes encoding RuBisCO in Sparkling Lake could not be classified, but we speculate that this expression may be originating from a nonphotosynthetic taxon. RuBisCO genes were also lower in rank order expression in Sparkling Lake than in the other two lakes. Also as expected, genes related to carboxylate transport (indicating likely use of algal exudates or abiotic photodegradation products as substrates) and rhodopsin biosynthesis were more highly expressed in the day time in all three lakes.

Our analysis of variance analysis showed that the high transcriptomic variability we observed is the product of biological variability rather than technical differences between samples. This underscores the dynamic nature of freshwater microbial communities, and the rapid rate at which microbes can alter their gene expression profiles in response to environmental changes. These results are in contrast to diel metatranscriptomic experiments performed in marine microbial communities, suggesting that smaller bodies of water that are closer to land experience higher levels of environmental variability, which in turn shapes the variability in microbial responses to changing conditions. This is an important factor in considering metatranscriptomic studies in freshwater systems, since multiple replicates, longer time periods, and more sampling overall may be required to disentangle fine-scale differences between environments.

Given the variability present in our datasets, we focused on differences that could be observed between day and night samples. While many genes had differential expression in day versus night, we chose to focus on heterotrophic carbon transport, as transporter expression has previously been used in marine systems to predict substrate use [(30)](https://paperpile.com/c/kQRinB/KlQV). Genes related to sugar transport were more highly expressed at night, regardless of lake or the type of sugar transported. Fewer types of sugar transporters were expressed in Sparkling Lake, perhaps relating to the observed disconnect between expression of genes contributing to primary production. The types of sugar transporters expressed are consistent with the composition of known freshwater algal exudates [(5)](https://paperpile.com/c/kQRinB/ryBs). This suggests that sugars may be important metabolites exchanged between phototrophs and heterotrophs in pelagic freshwater systems.

Sugar degradation at night is remarkably similar to how these metabolisms are partitioned in plant and algae cells: photosynthesis fuels carbon fixation in chloroplasts during the day, storing fixed carbon as starch, which is then broken into carbohydrates and used to fuel respiration by mitochondria at night [(31, 32)](https://paperpile.com/c/kQRinB/EB48+kRwL). We searched for genes encoding starch synthase and starch phosphorylase in our metatranscriptomic dataset, but we found that these genes had low expression and were primarily derived from Bacteroidetes and Cyanobacteria. This does not necessarily mean that diel starch accumulation and degradation is not occurring in our lakes. Rather, little to no expression of these genes may be due to the observed disconnect between expression levels and protein abundance [(33)](https://paperpile.com/c/kQRinB/oPQ7) or to the lack of representative algal genomes from our lakes. Still, this suggests a level of organization in microbial communities is analogous to that observed in eukaryotic cells.

There has been debate about whether actively growing and healthy phototrophic microbes exude carbon or whether these compounds are derived from decaying algae [(34)](https://paperpile.com/c/kQRinB/UiGK). Algae release cytotoxins that would presumably improve fitness by reducing competition [(35)](https://paperpile.com/c/kQRinB/Xfct), making it unlikely that these compounds are decay products. The marine phototroph *Prochlorococcus* likely exudates carbon to maintain redox, as it generates more reducing power than it can use via photosynthesis [(36)](https://paperpile.com/c/kQRinB/S7Dz). However, a frequently observed adaptation to excess reducing power is to downregulate photosynthesis electron flux, which is not observed in *Prochlorococcus.* In marine communities, heterotrophic bacterioplankton are highly dependent upon these *Prochlorococcus* exudates and likely performs a critical community function in return, such as the detoxification of hydrogen peroxide or free radicals (Morris et al, 2011, PLOS ONE).. 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 [(37)](https://paperpile.com/c/kQRinB/fZ0b).

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 interesting to note that the dependency between phototrophs and heterotrophs and the diel partitioning of carbon fixation and respiration would be anologous 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 that similar metabolic patterns predominate in microbial communities residing in three distinct freshwater lakes. Given the disparity in the biogeochemical features of these lakes, this provides evidence for conserved metabolic activities in freshwater ecosystems. Moreover, together with previous work in other aquatic systems, this highlights the presence if conserved metabolic exchanges that take place between photoautotrophs and heterotrophs that likely represent key organization principles in these ecosystems. This dataset also tells us about the daily variability of metatranscriptomes in freshwater and the depth of sequencing required to observe temporal trends, which is crucial knowledge when planning future experiments to assess seasonal or regional trends in gene expression.

## References

1. [Coyte KZ, Schluter J, Foster KR. 2015. The ecology of the microbiome: Networks, competition, and stability. Science 350:663–666.](http://paperpile.com/b/kQRinB/i0HM)

2. [Cole JJ. 1982. Interactions Between Bacteria and Algae in Aquatic Ecosystems. Annu Rev Ecol Syst 13:291–314.](http://paperpile.com/b/kQRinB/nZXs)

3. [Paver SF, Hayek KR, Gano KA, Fagen JR, Brown CT, Davis-Richardson AG, Crabb DB, Rosario-Passapera R, Giongo A, Triplett EW, Kent AD. 2013. Interactions between specific phytoplankton and bacteria affect lake bacterial community succession. Environ Microbiol 15:2489–2504.](http://paperpile.com/b/kQRinB/4WjX)

4. [Paver SF, Kent AD. 2010. Temporal patterns in glycolate-utilizing bacterial community composition correlate with phytoplankton population dynamics in humic lakes. Microb Ecol 60:406–418.](http://paperpile.com/b/kQRinB/O7uE)

5. [Giroldo D, Vieira AAH. 2005. Polymeric and free sugars released by three phytoplanktonic species from a freshwater tropical eutrophic reservoir. J Plankton Res 27:695–705.](http://paperpile.com/b/kQRinB/ryBs)

6. [Hellebust JA. 1965. EXCRETION OF SOME ORGANIC COMPOUNDS BY MARINE PHYTOPLANKTON1. Limnol Oceanogr 10:192–206.](http://paperpile.com/b/kQRinB/jrY2)

7. [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. Appl Environ Microbiol 77:7307–7315.](http://paperpile.com/b/kQRinB/usa5)

8. [Goldstone JV, Pullin MJ, Bertilsson S, Voelker BM. 2002. Reactions of hydroxyl radical with humic substances: bleaching, mineralization, and production of bioavailable carbon substrates. Environ Sci Technol 36:364–372.](http://paperpile.com/b/kQRinB/Ucg0)

9. [Dwulit-Smith JR, Hamilton JJ, Stevenson DM, He S, Oyserman BO, Moya-Flores F, Amador-Noguez D, McMahon K, Forest KT. 2018. acI Actinobacteria Assemble a Functional Actinorhodopsin with Natively-synthesized Retinal.](http://paperpile.com/b/kQRinB/IjYF)

10. [Vila-Costa M, Sharma S, Moran MA, Casamayor EO. 2013. Diel gene expression profiles of a phosphorus limited mountain lake using metatranscriptomics. Environ Microbiol 15:1190–1203.](http://paperpile.com/b/kQRinB/R3gI)

11. [Poretsky RS, Hewson I, Sun S, Allen AE, Zehr JP, Moran MA. 2009. Comparative day/night metatranscriptomic analysis of microbial communities in the North Pacific subtropical gyre. Environ Microbiol 11:1358–1375.](http://paperpile.com/b/kQRinB/32NM)

12. [Aylward FO, Eppley JM, Smith JM, Chavez FP, Scholin CA, DeLong EF. 2015. Microbial community transcriptional networks are conserved in three domains at ocean basin scales. Proc Natl Acad Sci U S A 112:5443–5448.](http://paperpile.com/b/kQRinB/l3sB)

13. [Bendall ML, Stevens SL, Chan L-K, Malfatti S, Schwientek P, Tremblay J, Schackwitz W, Martin J, Pati A, Bushnell B, Froula J, Kang D, Tringe SG, Bertilsson S, Moran MA, Shade A, Newton RJ, McMahon KD, Malmstrom RR. 2016. Genome-wide selective sweeps and gene-specific sweeps in natural bacterial populations. ISME J 10:1589–1601.](http://paperpile.com/b/kQRinB/bx1Z)

14. [Garcia SL, Stevens SLR, Crary B, Martinez-Garcia M, Stepanauskas R, Woyke T, Tringe SG, Andersson SGE, Bertilsson S, Malmstrom RR, McMahon KD. 2018. Contrasting patterns of genome-level diversity across distinct co-occurring bacterial populations. ISME J 12:742–755.](http://paperpile.com/b/kQRinB/I0sv)

15. [Satinsky BM, Gifford SM, Crump BC, Moran MA. 2013. Use of internal standards for quantitative metatranscriptome and metagenome analysis. Methods Enzymol 531:237–250.](http://paperpile.com/b/kQRinB/EXys)

16. [Nielsen HB, Almeida M, Juncker AS, Rasmussen S, Li J, Sunagawa S, Plichta DR, Gautier L, Pedersen AG, Le Chatelier E, Pelletier E, Bonde I, Nielsen T, Manichanh C, Arumugam M, Batto J-M, Quintanilha Dos Santos MB, Blom N, Borruel N, Burgdorf KS, Boumezbeur F, Casellas F, Doré J, Dworzynski P, Guarner F, Hansen T, Hildebrand F, Kaas RS, Kennedy S, Kristiansen K, Kultima JR, Léonard P, Levenez F, Lund O, Moumen B, Le Paslier D, Pons N, Pedersen O, Prifti E, Qin J, Raes J, Sørensen S, Tap J, Tims S, Ussery DW, Yamada T, MetaHIT Consortium, Renault P, Sicheritz-Ponten T, Bork P, Wang J, Brunak S, Ehrlich SD, MetaHIT Consortium. 2014. Identification and assembly of genomes and genetic elements in complex metagenomic samples without using reference genomes. Nat Biotechnol 32:822–828.](http://paperpile.com/b/kQRinB/ZxdR)

17. [2014. BBMap: A Fast, Accurate, Splice-Aware Aligner.](http://paperpile.com/b/kQRinB/I8ZD)

18. [Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Prjibelski AD, Pyshkin AV, Sirotkin AV, Vyahhi N, Tesler G, Alekseyev MA, Pevzner PA. 2012. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J Comput Biol 19:455–477.](http://paperpile.com/b/kQRinB/D8P2)

19. [Kopylova E, Noé L, Touzet H. 2012. SortMeRNA: fast and accurate filtering of ribosomal RNAs in metatranscriptomic data. Bioinformatics 28:3211–3217.](http://paperpile.com/b/kQRinB/kbRD)

20. [Pruitt KD, Maglott DR. 2001. RefSeq and LocusLink: NCBI gene-centered resources. Nucleic Acids Res 29:137–140.](http://paperpile.com/b/kQRinB/fX0s)

21. [Huang Y, Niu B, Gao Y, Fu L, Li W. 2010. CD-HIT Suite: a web server for clustering and comparing biological sequences. Bioinformatics 26:680–682.](http://paperpile.com/b/kQRinB/Pc8N)

22. [Liao Y, Smyth GK, Shi W. 2014. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. Bioinformatics 30:923–930.](http://paperpile.com/b/kQRinB/Rit1)

23. [Kang DD, Froula J, Egan R, Wang Z. 2015. MetaBAT, an efficient tool for accurately reconstructing single genomes from complex microbial communities. PeerJ 3:e1165.](http://paperpile.com/b/kQRinB/taMr)

24. [Parks DH, Imelfort M, Skennerton CT, Hugenholtz P, Tyson GW. 2015. CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. Genome Res 25:1043–1055.](http://paperpile.com/b/kQRinB/1ntJ)

25. [Markowitz VM. 2006. The integrated microbial genomes (IMG) system. Nucleic Acids Res 34:D344–D348.](http://paperpile.com/b/kQRinB/fhFA)

26. [Yang E-W, Girke T, Jiang T. 2013. Differential gene expression analysis using coexpression and RNA-Seq data. Bioinformatics 29:2153–2161.](http://paperpile.com/b/kQRinB/VqZR)

27. [Rosenberg E. 2014. The Family Chitinophagaceae, p. 493–495. *In* The Prokaryotes.](http://paperpile.com/b/kQRinB/mk5V)

28. [Wilkinson CR. 1979. Bdellovibrio-like parasite of cyanobacteria symbiotic in marine sponges. Arch Microbiol 123:101–103.](http://paperpile.com/b/kQRinB/99Cl)

29. [Cavalier-Smith T. 2006. Rooting the tree of life by transition analyses. Biol Direct 1:19.](http://paperpile.com/b/kQRinB/5KOP)

30. [Poretsky RS, Sun S, Mou X, Moran MA. 2010. Transporter genes expressed by coastal bacterioplankton in response to dissolved organic carbon. Environ Microbiol 12:616–627.](http://paperpile.com/b/kQRinB/KlQV)

31. [Graf A, Schlereth A, Stitt M, Smith AM. 2010. Circadian control of carbohydrate availability for growth in Arabidopsis plants at night. Proc Natl Acad Sci U S A 107:9458–9463.](http://paperpile.com/b/kQRinB/EB48)

32. [Vitova M, Bisova K, Kawano S, Zachleder V. 2015. Accumulation of energy reserves in algae: From cell cycles to biotechnological applications. Biotechnol Adv 33:1204–1218.](http://paperpile.com/b/kQRinB/kRwL)

33. [Moran MA, Satinsky B, Gifford SM, Luo H, Rivers A, Chan L-K, Meng J, Durham BP, Shen C, Varaljay VA, Smith CB, Yager PL, Hopkinson BM. 2013. Sizing up metatranscriptomics. ISME J 7:237–243.](http://paperpile.com/b/kQRinB/oPQ7)

34. [Sharp JH. 1977. Excretion of organic matter by marine phytoplankton: Do healthy cells do it?1. Limnol Oceanogr 22:381–399.](http://paperpile.com/b/kQRinB/UiGK)

35. [Kind T, Meissen JK, Yang D, Nocito F, Vaniya A, Cheng Y-S, Vandergheynst JS, Fiehn O. 2012. Qualitative analysis of algal secretions with multiple mass spectrometric platforms. J Chromatogr A 1244:139–147.](http://paperpile.com/b/kQRinB/Xfct)

36. [Braakman R, Follows MJ, Chisholm SW. 2017. Metabolic evolution and the self-organization of ecosystems. Proc Natl Acad Sci U S A 114:E3091–E3100.](http://paperpile.com/b/kQRinB/S7Dz)

37. [Nelson CE, Goldberg SJ, Wegley Kelly L, Haas AF, Smith JE, Rohwer F, Carlson CA. 2013. Coral and macroalgal exudates vary in neutral sugar composition and differentially enrich reef bacterioplankton lineages. ISME J 7:962–979.](http://paperpile.com/b/kQRinB/fZ0b)