Metabolic Shifts in Marine Phytoplankton From Viral Infection and Diel Cycles Uncovered Using Dynamic Bayesian Networks

Xiang Liu
University of Arizona https://orcid.org/0000-0002-7081-3694
Bonnie L. Hurwitz (bhurwitz@email.arizona.edu)
University of Arizona https://orcid.org/0000-0001-8699-957X

Article

Keywords: DBN, Metabolic, marine

Posted Date: September 29th, 2020

DOI: https://doi.org/10.21203/rs.3.rs-54982/v1

Abstract

Viruses play a fundamental role in the ecological dynamics of marine microbes by influencing host metabolic flux and altering population size via viral lysis. Despite recent advances in understanding the temporal and environmental factors that drive community-level microbial interactions, viral activity is currently unaccounted for in biogeochemical models. Moreover, the host's metabolic response to viral infection in natural systems remains elusive. Here, we examine changes in metabolism in marine microbial communities related to virus activity and diel cycles in the surface oceans in the North Pacific Subtropical Gyre (NPSG). We use a time-varying dynamic Bayesian network (DBN) to untangle gene co-expression networks in dominant microbes associated with the environment. The resulting data suggest that viruses play a significant role in driving microbial metabolic flux on par with diel cycles. Most notably, viruses play an opportunistic role by synchronizing viral lysis with diel cycles in their photoautotrophic hosts; and continually infecting heterotrophic hosts by co-opting their alternative energy production strategies to drive replication throughout the day. This work provides a unified model that accounts for both viral activity and diel cycles to predict the distribution, variability, and trajectory of dominant microbes and their key metabolic pathways in global ocean processes.

Introduction

Viruses infect hosts from all domains of life and are the most abundant biological entities on Earth with more than 10^{31-42} particles [1]. The vast majority of viruses are phages that infect bacteria and archaea [1]. Viral populations vary based on host abundance that fluctuates given physical and chemical properties associated with the environment [2–4]. For example, marine cyanophage take advantage of their host's diel replication strategies by co-opting photosynthetic energy production during the day to drive viral infection and lysis at dusk [5]. Viruses also encode and express host-derived genes called Auxiliary Metabolic Genes (AMGs) [6] that alter host metabolism during viral infection [7]. Although recent work has elucidated the temporal dynamics of viral populations in natural systems [5], the host's metabolic response to daily environmental oscillations combined with viral infection remains elusive.

In the euphotic zone of the ocean, marine phytoplankton exhibit diel patterns in the transcriptional activity that optimize photosynthesis and nutrient uptake [8–18]. In particular, light and nutrient-limitation in the open ocean drive oscillations in critical metabolic processes in bacteria. Recent work has shown that phototrophic and heterotrophic bacteria alternate their use of limited nutrients such as nitrogen and phosphorus according to diel cycles to optimize the aggregate community metabolism [19]. Moreover, primary production via phototrophic bacteria peaks in the afternoon [19], followed by viral activity and lysis at dusk [5]. Heterotrophic bacteria take in sugar/lipids and generate energy via the TCA cycle at night [19]. These recent analyses of diel rhythms in the ocean have accounted for functional shifts in bacteria, eukaryotes, and viruses independently. Yet, none to date examine the cumulative effects of these processes in a single model.

To examine metabolic shifts and temporal dynamics in microbial communities associated with virus activity, we leveraged a large-scale metagenomic and metatranscriptomic time-series from viral and bacterial size-fractionated water samples from the open ocean in the North Pacific Subtropical Gyre (NPSG) [5,10]. Specifically, we used a time-varying dynamic Bayesian network (DBN) inference to find functional shifts in microbial populations associated with diel patterns and virus activity from an 8-day cruise with surface water sampled over 4-hour intervals. Because virus activity is tightly coupled to the host and their environment, we constructed a DBN to detect dependencies between gene modules (co-expressed genes), virus activity, and environmental parameters from dominant photoautotrophs and heterotrophs in the ocean. These viral-host metabolic interactions offer essential clues into host-specific adaptations, nutritional constraints, and metabolic bottlenecks, while also accounting for the effects of viral infection. Our work provides a comprehensive view of species-specific metabolism over diel cycles with virus infection for a holistic view of the functional role of dominant microbes in the ocean.

Methods

Sampling and experimental design

We leveraged existing metagenomics and metatranscriptomics data from a cruise in the North Pacific Subtropical Gyre (NPSG) taken during the Hawaii Ocean Experiment Legacy II cruise (KM1513) from July 25th and August 5th, 2015. As previously described [5,11], samples were collected every four hours (2:00, 6:00, 10:00, 14:00, 18:00 and 22:00) during two periods of diel measurements within the same water mass with a maximum depth of 15 m, yielding samples from a total of 44 time-points. For each time-point, three samples were generated corresponding to two cellular fractions (> 0.2 μm) and one viral fraction (0.2 μm > 0.03 μm). General cruise information and associated biogeochemical and oceanographic measurements are available online [20]. Sample metadata used in this study can be found in Supplementary Dataset 1. Our workflow of the following analysis is shown in Supplementary Figure 1

Metagenome and metatranscriptome sequencing

Detailed information about sequencing methods was described in a previous study [11]. Briefly, metagenomes were created from the viral fractions and metagenomes and transcriptomes from the cellular fractions. Metagenomic library preparation was performed using the Illumina TruSeq Nano LT library preparation kit for the cellular fraction. Illumina Neoprep library automation instrument and a Neoprep compatible TruSeq Nano LT kit were used for metagenome library preparation of viral fraction (Illumina). Metatranscriptomic samples were prepared using 5-50 ng of total RNA to the ScriptSeq cDNA V2 library preparation kit (Epicentre). Quantitative standards for transcriptomes were also spiked-in after extraction but before library preparation, consistent with previous methods [21]. Illumina Nextseq500 system was used to sequencing both metagenomes and metatranscriptomes. Raw metagenomic and metatranscriptomic reads were stored in NCBI Sequence Read Archive (SRA) under BioProject accession PRJNA358725.

Generation of viral scaffold metatranscriptome abundance dataset

A normalized metatranscriptome abundance dataset of viral scaffolds was obtained from a previous paper for this study [5]. Briefly, all metagenomes (total of 88 samples) were assembled individually using Mira v. 4.9.5_2 [22]. The resulting assembled contigs were analyzed by VirSorter [23] to extract viral sequences. Those viral sequences were re-assembled and then analyzed by VirSorter a second time, with those contigs annotated in categories 1 and 2 retained as putative viral contigs. SSPACE 3.0 [24] was used for scaffolding. Metatranscriptomes (total of 44 samples) were analyzed with an end-joining, quality-trimming, read-mapping, and quantitative normalization workflow, as described in a previous study [5]. Based on the viral annotation files [5], viral scaffolds identified with their host of *Prochlorococcus marinus* and *Candidatus Pelagibacterubique* were extracted from the final normalized transcriptome abundance dataset for downstream analysis in this study.

Generation of bacterial metatranscriptome abundance dataset

For the bacterial metatranscriptome, raw reads were trimmed with trim_galore (default setting) [25], then reads of each sample were mapped to *Prochlorococcusmarinus* and *Candidatus Pelagibacterubique* contigs identified from the Station ALOHA Gene Catalogue V1 (available online: https://www.imicrobe.us/#/projects/263) with Bowtie2 [26]. The data were then processed in two steps before generating the final metatranscriptome matrix. First, low-abundance genes for each sample were removed with a cutoff of 15, resulting in a total of 11,010 unique genes for *Prochlorococcus marinus* and 4,473 unique genes for *Candidatus Pelagibacterubique*. Second, genes that appeared in less than 20% of the sample size were removed to reduce the dataset's sparseness. This process yielded the final metatranscriptome matrix with 1829 unique genes for *Prochlorococcus marinus* and 2510 unique genes for *Candidatus Pelagibacterubique.*

Identifying gene modules

We used weighted gene co-expression network analysis (WGCNA) to group genes into gene modules based on their co-expression patterns [27]. The pre-processed datasets were normalized using methods with the "varianceStabilizingTransformation" command from the DESeq2 package [28]. Normalized counts were analyzed using the R packages WGCNA [27]. For each dataset, the smallest soft threshold that would achieve an R^2 value > 0.8 for fit to a scale-free topology was chosen from the "pickSoftThreshold" command in WGCNA. We selected a power of 6 and 5 for *Prochlorococcus marinus* and *Candidatus Pelagibacterubique* datasets, respectively (Supplementary Figure 2). Next, adjacency matrices were calculated based on the soft power chosen from the last step. Based on adjacency matrices, a Topology Overlap Matrix (TOM) was created by the "TOMsimilarity" command. Gene modules were identified using the "cutreeDynamic" command in WGCNA with a minimum module size of 30. Finally, the module "eigengenes" [29] for each module were calculated using the "moduleEigenegenes" command in WGCNA. Gene modules with similar eigengenes were merged. Genes without any module assignments because of little correlation with any other genes were designated as module 0 (Supplementary Figure 3 and

Supplementary Figure 4). The final dataset consists of 5 and 10 gene modules (exclude module 0) for *Prochlorococcus marinus* (Supplementary Dataset 2) and *Candidatus Pelagibacterubique* (Supplementary Dataset 3), respectively.

Bayesian network modeling

Datasets for each species (*Prochlorococcus marinus* and *Candidatus Pelagibacterubique)* were divided into day (including samples collected at 10:00, 14:00 and 18:00) and night (including samples collected at 22:00, 2:00, and 6:00) groups, yielding a total of 4 datasets. Viral activity and surface light intensity (surface PAR) were added to build BNs. The viral activity was calculated as the total sum of normalized counts within each sample. Gene counts were aggregated by the gene modules, and Total Count (TC) was used to normalize the datasets.

In this study, we built BNs for each species and combined the two species into a single BN. We use R [30] and the bnlearn package [31] to perform Bayesian network structure learning, parameter learning, and model validation. A two-stage Dynamic Bayesian Network (2TDBN) with Tabu search algorithm was used. Also, to find the simplest network, we compared the performance under different settings of the maximum number of parents (MNP: from 1 to 10). We ran a 10-fold cross-validation ten times to validate all the modeling strategies. The Bayesian information criterion (BIC) score was used for model selection. An averaged network structure is created by Cytoscape [32] to show a constant network; where the edges are weighted by their frequency from the cross-validation process (threshold = 0.6). Because 2TDBN only considers “current time” samples and their immediately previous time samples, each sample was paired with the previous sample. This model limited edges to only those going from prior time points (T_0) to subsequent time points (T_1) for the same variable. In other words, the value of a variable at T_1 is only affected by its value at T_0 or values of other variables at T_1. The function to calculate the BIC score is shown below:

$$BIC = \log(L) - k * nparams(x) \quad (1)$$

Where,

L: is the likelihood function

nparams(x): is the number of parameters in a network x

k: is the penalty term

Because our variables are all continuous, we assume the Bayesian network is a Gaussian Bayesian Network (GBN) that all variables are Gaussian and all conditional probability distributions (CPDs) are linear Gaussian [33]. Let Y has a linear Gaussian of its parents $\boldsymbol{X}$, with parameters β_0, $\boldsymbol{\beta}$, and σ^2, then it has:

$$P(Y|\boldsymbol{X}) \sim N(\beta_0 + \beta^T X; \sigma^2) \quad (2)$$

Results

Our strategy for examining species-specific transcription over diel cycles while accounting for viral activity is shown in Supplementary Fig. 1. To begin, we retrieved metatranscriptomics reads from 44-time points taken during a cruise in the North Pacific Subtropical Gyre (NPSG) from July 25th and August 5th, 2015. Because the transcripts were derived from the entire bacterial community, and our focus is on species-specific metabolic processes with viral activity, we mapped the reads to individual species in the Station ALOHA Gene Catalogue. To compare and contrast lifestyles and metabolic capacity in natural populations that shift with diel cycles, we selected genes from 1-dominant photoautotroph (*Prochlorococcus Marinus*) and 1-dominant heterotroph (*Candidatus Pelagibacter ubique)*. These data were used to create transcriptome abundance tables for each species that were used in subsequent network analyses to find genes that are co-expressed (Supplementary Dataset 2 and Supplementary Dataset 3).

The resulting gene modules (co-expressed genes) were then linked to diel cycles and viral activity using Dynamic Bayesian Network analyses (Fig. 1). Data on viral activity for *Prochlorococcus Marinus* or *Candidatus Pelagibacter ubique* was derived from prior work [5] by summing up the normalized genes counts for viruses linked to those hosts. Data on diel cycles was derived from surface photosynthetic active radiation (PAR) values taken during sampling. This novel pipeline allowed us to examine species-specific metabolic functions, in conjunction with viral activity, that may be overlooked when combined in a community-level analysis. Moreover, by carefully separating the metabolic activity of photoautotrophic and heterotrophic populations, we can examine their synchronized metabolic activity, in natural communities, and with viral activity.

Diel patterns in transcription activity for dominant microbes in the euphotic zone. To examine species-specific gene modules in the metatranscriptomic data, we distilled down the dataset to two dominant species (1-photoautotroph and 1-heterotroph) to explore fine-scale patterns in gene co-expression related to diel cycles and viral activity. Specifically, we used weighted gene co-expression network analysis (WGCNA) to group genes into gene modules based on their co-expression patterns for *Prochlorococcus Marinus* (a photoautotroph, Fig. 1a) and *Candidatus Pelagibacter ubique* (a heterotroph, Fig. 1b). The complete list of KEGG annotated for each module were shown in Supplementary Dataset 4 and Supplementary Dataset 5. Overall, 3 gene modules for *Prochlorococcus Marinus* showed diel patterns (during the day, night/dawn, dusk/night, and dusk) 4 for *Candidatus Pelagibacter ubique* (during the day, day/dusk, and night; see Supplementary Table 1). Each of these gene modules is described in detail below.

Bayesian network analysis and dynamic dependencies on diel cycles and viral activity.

Next, we examined temporal and environmental dynamics in gene modules for each species (*Prochlorococcus Marinus and Ca. Pelagibacter ubique*) using a Dynamic Bayesian Network (DBN) analysis. To gain insight into complex interactions between each species and the environment, we

examined the dynamic dependencies of surface PAR and viral activity in the network. Because both species are key contributors to microbial dynamics at Station ALOHA, known interactions between taxa and genes can be used to validate our models. Moreover, the data can be used to build predictive models for dominant photoautotrophs and heterotrophs based on previous time-points, and potentially uncover new biological interactions related to diel cycles or viral activity.

To examine biological interactions related to diel cycles and viral activity in *Prochlorococcus Marinus* and *Ca. Pelagibacter ubique*, we obtained an averaged network structure for each species independently (Fig. 2a & b) and for the two species combined (Fig. 2c). The resulting networks contained nodes that are either gene modules, or environmental parameters (viral activity or surface PAR). Lines connecting each of the nodes indicate directed links between nodes at different timepoints, wherein the width of the line indicates the strength of the bootstrap support. For example, nodes that are strongly connected to viral activity (strength = 1; red) or surface PAR (green) have 10 times 10-fold cross-validation, meaning of one hundred networks 100% of the networks show this connection (Fig. 2).

Diel Patterns in Prochlorococcus Marinus **and** Ca. Pelagibacter ubique

Prochlorococcus Marinus has increased transcription of genes related to photosynthesis and protein folding during the day. We found a significant connection between *p_module_1* for *Prochlorococcus Marinus* and Surface PAR in the DBN (Fig. 2a-day; strength = 1). This module is over-expressed during the day (Fig. 1a) and contains genes that are related to photosynthesis and protein biogenesis, as described below. *p_module_1* for *Prochlorococcus Marinus* showed increased transcription during the day for genes related to chaperones and protein folding (*clpC, ftsH, hflB*), chlorophyll metabolism (*chlH, chlE*), RNA degradation (*rnj*), and photosynthesis (*psbA, psbC, psaA*; Supplementary Table 1). Prior work shows that photoautotrophs have increased transcription for genes related to light-capture and protein synthesis with increased surface PAR [9–11], which is consistent with these findings. *clpC* has been previously shown to play an important role in oxidative stress [34], and could be important in preventing photooxidative stress in biological molecules such as proteins, lipids, and DNA, or more specifically photosystem components. Interestingly, we also identified an uncharacterized putative membrane protein (K06890) in *p_module_1* that is associated with COG0670 (*YbhL*), a gene that is known to interact with *ftsH*. These membrane proteins are essential for disassembly and oligomerization of protein complexes during protein biogenesis [35]. These findings suggest that *Prochlorococcus Marinus* has increased levels of transcription in genes related to photosynthesis, removal of oxygen free radicals that can harm photosystem components, and chlorophyll metabolism.

Prochlorococcus Marinus has increased transcription of genes related to oxidative phosphorylation and protein synthesis at night. At night/dawn, *Prochlorococcus Marinus* shows increased transcription in *p_module_3* for oxidative phosphorylation (*atpA*), and ribosome and protein synthesis (*pnp, rplB, rpsC, tuf;* Supplementary Table 1). These data support the hypothesis that cyanobacteria alternate between metabolic processes for photosynthesis during the day and protein synthesis during the night [19].

Ca. Pelagibacter ubique prevents protein damage, regulates transcription, and scavenges sulfur from DMSP **during the day**. *Ca. Pelagibacter ubique* is the most abundant heterotrophic bacteria in the ocean. In the daytime, this species contains abundant transcripts in *cp_module_1* (day/dusk), *cp_module_2* (day), and *cp_module_4* (day) that contain genes for RNA degradation and biosynthesis (*groEL). cp_module_2* and *cp_module_4* are associated with RNA degradation, chaperones and folding (*dnaK, ftsH;* Supplementary Table 1). In the DBN, *cp_module_1* and *cp_module_4* both showed strong connections (strength = 1) with surface PAR in *Ca. Pelagibacter ubique* (Fig. 2b-day). Previous studies on *Ca. Pelagibacter ubique* indicates that *groEL, dnaK*, and *ftsH* are among the most abundant proteins in marine pelagic populations [36]. Interestingly, these genes can assist in protein folding and proteolysis to prevent protein damage and exposure to environmental stresses [36]. *cp_module_2* also contains abundant transcripts for *S*-Adenosyl-l-homocysteine (*ahcY)*. The resulting product of *ahcY is S*-adenosyl-l-methionine (SAM) methyltransferases [37], which is an important riboswitch in *Ca*. *Pelagibacter ubique* and regulates transcription [38]. *ahcY* transcripts are co-expressed with the ABC transporter gene (*proX*) in *cp_module_1* and *cp_module_2*. *proX* is involved in glycine betaine and proline betaine uptake systems [39] and allows marine bacteria to degrade dimethylsulfoniopropionate (DMSP) [40–42] to create volatile sulfur species such as dimethylsulfide (DMS). Thus, the upregulation of *proX* during the day may allow *Ca. Pelagibacter ubique* to scavenge reduced sulfur from DMSP [43–45] for cell growth, which results in the production of DMS for cloud production and cooling. This species also appears to use gluconeogenesis during the day (*aldB* in *cp_module_1*) to breakdown pyruvate its primary carbon source [46], and generate energy via the TCA cycle (*sdhA* in *cp_module_1*). Finally, *cp_module_2* and *cp_module_4* contain genes related to nitrogen uptake systems including ammonium transporter *(amtB)* and the branched amino acid transport system *(livK)* that could help to drive the growth in N-limiting conditions in the open ocean.

Ca. Pelagibacter ubique transcribes genes for energy production at night. *cp_module_6* in *Ca. Pelagibacter ubique* shows increased transcription overnight related to oxidative phosphorylation (*atpA, atpD*), glycolysis and gluconeogenesis (*aldB*), and elongation factors (*tuf*; Supplementary Table 1). These findings suggest that ATP production by oxidative phosphorylation occurs at night and may play a role in breaking down compounds released by phytoplankton when they bloom and are lysed by viruses during the day. These findings are consistent with prior work that shows that light does not influence growth rates in *Ca*. Pelagibacter ubique [47]. Thus, *Ca*. Pelagibacter ubique replicates throughout the day by generating energy from 1) proteorhodopsin energy production via proton motives forces across the cell membrane during the day [48] and 2) increased oxidative phosphorylation due to carbon and nutrient breakdown at night (*cp_module_6*, Fig. 1b).

Viral Activity in Prochlorococcus Marinus **and** Ca. Pelagibacter ubique

Viral infection in Prochlorococcus Marinus occurs at dusk redirecting host energy, nucleotide, and carbon metabolism. In *Prochlorococcus Marinus, p_module_6* is strongly connected to viral activity in the DBN (Fig. 2a-day; Supplementary Table 2) and contains genes that are known to be overexpressed in the host during viral infection [49]. Genes in *p_module_6* are overexpressed at dusk (Fig. 1a), which is consistent

with prior studies of temporal viral activity [5]. Interestingly, viral activity has a strong connection to itself during the day (Fig. 2a-day), and none of the gene modules have a significant association with viral activity at night (Fig. 2a-night), indicating that viral activity is highly time-dependent. *p_module_6* contains genes related to carbon metabolism (*talA*), DNA replication and repair (*recA*), and nucleotide metabolism (*nrdJ*) that are encoded and expressed by cyanophage during infection [49]. Previously, *talA* was shown to be the most highly expressed transcript in cultured marine cyanobacteria during viral infection. Phage gene expression of *talA* is thought to redirect carbon from the Calvin cycle in the host to the Pentose Phosphate Pathway (PPP) for viral energy production [49]. Interestingly this metabolic shift simultaneously creates reducing power (via the PPP) and the carbon skeleton needed for nucleotide metabolism, which is a limited resource for phage production [7,49]. Further, *p_module_6* contains *nrdJ* that may be overexpressed for nucleotide metabolism for phage production [7]. Similarly, *recA* is thought to play a role in phage replication by repairing DNA damage from UV damage radiation [5]. *p_module_6* also contains genes related to hyper-modification (*glyA*) that may protect bacteriophage DNA and prevent degradation by the host (*clpX)* [50]. These opposing genes may be involved in the molecular arms race between the phage and their host during infection. Finally, *p_module_6* contains *gltS*, a GS-GOGAT component, for nitrogen assimilation that was previously shown to peak in *Prochlorococcus* at dusk [5]. Phage particles are composed of up to 41% extracellularly derived N [51]. Therefore transcription of this gene at dusk may be important for viral replication and not host-nitrogen assimilation as previously suggested [19]. All in all, module 6 contains genes that may be over transcribed due to a viral take-over of host metabolism, rather than host-driven processes.

In comparison, *p_module_3* shows a weak association with viral activity during the day (Fig. 2a-day; strength = 0.71). *p_module_3* contains genes for oxidative phosphorylation and ribosome and protein synthesis that are overexpressed at night (Fig. 1a) and may be important for host protein synthesis [52]. The different network structures between day and night indicate that viral activity is driven by diel cycles in their photoautotrophic host, as previously described [5].

Gene modules linked to viral infection in Ca. Pelagibacter ubique. In *Ca. Pelagibacter ubique*, viral activity is strongly connected with *cp_module_4* and *cp_module_5* during the day (Fig. 2b-day, with strength equal to 1 and 0.87, respectively), compared to *cp_module_4* and *cp_module_8* at night (Fig. 2b-night; Supplementary Table 2). These data indicate that viruses may have different strategies for infection during the day vs night based on changes in the metabolic processes of their hosts (Fig. 2b). Therefore, although the host does not exhibit differences in growth rates due to diel cycles [47], viral infection strategies change. Abundant genes in *cp_module_4* are grouped in three categories: 1) RNA degradation and biogenesis (*groEL* and *dnaK*); 2) peptidases and inhibitors (*ftsH, hflB* and *hflK*); and 3) ammonia transporters (*amtB*). These genes are also found in *cp_module_2* and are overexpressed during the day (Fig. 1b) but not linked to viral activity (Fig. 2b-day). Given that *dnaK, groEL*, and *ftsH* are abundant both in this host and marine pelagic populations [36], they could be expressed by the host in *cp_module_2* during the day and by phage in *cp_module_4* during the day and night to bolster host fitness during

infection. *cp_module_5* includes genes related to glycolysis, gluconeogenesis and pyruvate metabolism (*aldB* and *lldD*), TCA cycle (*sdhA* and *frdA*), DNA repair and recombination (*recA*), and RNA degradation (*dnaK* and *HSPA9*). Interestingly, pyruvate for gluconeogenesis is a primary carbon source for SAR11 clades [53] and may help to drive replication using proteorhodopsin energy production [54] during the day. At night, genes in *cp_module_8* are related to unknown proteins and aminoacyl-tRNA synthetases (*thrS*). Unknown genes in *cp_module_8* require further study to elucidate their role in host metabolism and potentially for viral replication.

A combined network structure elucidates the interaction of each species over diel cycles and with viral infection. Microbes in the ocean do not exist in isolation, and instead are part of complex interactions both between species and the environment. Thus, to mimic these interactions we combined the species together in a unified model. Overall, our combined network (Fig. 2c) mimicked the structure of the networks for individual species but clarified the strength of connections between nodes. In contrast to other findings that growth rates in *Ca. Pelagibacter ubique* are not tied to diel patterns [47]; we see that genes in *cp_module_4* have strong connections with surface PAR (strength = 1, Fig. 2c). Specifically, genes related to preventing protein damage, regulating transcription, and scavenging sulfur from DMSP may have diel periodicity in natural *Ca. Pelagibacter ubique* populations. Interestingly, this same module is also closely connected with viral activity in both the day and night. Compared to *Prochlorococcus Marinus* that has strong connections to viral activity during the day only. Viral diel patterns for *Prochlorococcus Marinus* are related to gene co-expression patterns in photosynthesis, energy production, and nucleotide biosynthesis (in *p_module_6*). Further, viral activity at night in *Ca. Pelagibacter ubique* is driven both by genes in *cp_module_4* and viral activity in *Prochlorococcus Marinus.* These data suggest that viral lysis of photoautotrophic bacteria helps to drive viral activity in heterotrophs like *Ca. Pelagibacter ubique* likely due to increased growth rates from available nutrients. Thus, the differences between the individual species and combined networks indicate that there is more information we can acquire when considering both species as a community. Moreover, viruses appear to be a major driver in microbial community dynamics, on par with diel cycles.

Discussion

Aylward and colleagues found that abundant photoautotrophs “set the whole-community diel cycle” in both the coastal and pelagic ocean [55]. Our data validate and extend this hypothesis to include a fundamental role for viruses in diel dynamics in the ocean. Although abundant photoautotrophs “set the stage” for diel cycles through increased primary production during the day, viruses shift the system at dusk by lysing photoautotrophs and making limited nutrients and carbon available for heterotrophic bacteria. Our unified Dynamic Bayesian Network validates these long-standing hypotheses and solidifies the role that viruses play, showing that viruses and light play complementary and primary roles in microbial population dynamics. Where prior studies have examined the transcriptome dynamics of individual species or microbial communities [55], our study takes a holistic view by including viruses in the dynamic interplay between hosts and the marine environment.

By intentionally distilling down the system to two well-studied dominant bacteria (*Prochlorococcus Marinus and Ca. Pelagibacter ubique*) we can both validate and extend our knowledge of species-species metabolic interactions in the ocean. Moreover, we can use this same platform to examine the role of viruses and sunlight using Dynamic Bayesian Networks (DBN). Our findings support metabolic studies in these individual species but extend the analysis to consider the weight of each factor in driving metabolic cycles in the entire community. Moreover, the gene co-expression analysis combined with a DBN gives us insight into which genes are expressed under which conditions. Our analysis pipeline offers a novel approach to detecting host metabolic responses to viral infection and sunlight in natural populations.

Declarations

Acknowledgments

This work was supported in part by Gordon and Betty Moore Foundation GBMF 8751 (to B.L.H) and Simons Foundation muSCOPE award ID 481471 (to B.L.H).

Competing Interests

The authors declare that they have no conflict of interest.

References

1. Breitbart, M. & Rohwer, F. Here a virus, there a virus, everywhere the same virus? *Trends Microbiol.***13**, 278–284 (2005).
2. Hurwitz, B. L., Brum, J. R., Sullivan, M. B. & Hurwitz, B. L. Depth-stratified functional and taxonomic niche specialization in the 'core'and 'flexible'1. *Virus***19**, 20.
3. Hurwitz, B. L., Westveld, A. H., Brum, J. R. & Sullivan, M. B. Modeling ecological drivers in marine viral communities using comparative metagenomics and network analyses. *PNAS***111**, 10714–10719 (2014).
4. Brum, J. R. *et al.* Patterns and ecological drivers of ocean viral communities. *Science***348**, (2015).
5. Aylward, F. O. *et al.* Diel cycling and long-term persistence of viruses in the ocean's euphotic zone. *Proc. Natl. Acad. Sci. U. S. A.***114**, 11446–11451 (2017).
6. Breitbart, M., Thompson, L. R., Suttle, C. A. & Sullivan, M. B. Exploring the Vast Diversity of Marine Viruses. *Oceanography***20**, 135–139 (2007).
7. Lindell, D. *et al.* Genome-wide expression dynamics of a marine virus and host reveal features of co-evolution. *Nature***449**, 83–86 (2007).
8. Vaulot, D., Marie, D., Olson, R. J. & Chisholm, S. W. Growth of Prochlorococcus, a photosynthetic prokaryote, in the equatorial Pacific Ocean. *Science***268**, 1480–1482 (1995).
9. Ottesen, E. A. *et al.* Pattern and synchrony of gene expression among sympatric marine microbial populations. *Proc. Natl. Acad. Sci. U. S. A.***110**, E488–97 (2013).

10. Ottesen, E. A. *et al.* Ocean microbes. Multispecies diel transcriptional oscillations in open ocean heterotrophic bacterial assemblages. *Science***345**, 207–212 (2014).
11. Wilson, S. T. *et al.* Coordinated regulation of growth, activity and transcription in natural populations of the unicellular nitrogen-fixing cyanobacterium Crocosphaera. *Nat Microbiol***2**, 17118 (2017).
12. Frischkorn, K. R., Haley, S. T. & Dyhrman, S. T. Coordinated gene expression between Trichodesmium and its microbiome over day–night cycles in the North Pacific Subtropical Gyre. *ISME J.***12**, 997–1007 (2018).
13. Becker, K. W. *et al.* Daily changes in phytoplankton lipidomes reveal mechanisms of energy storage in the open ocean. *Nat. Commun.***9**, 5179 (2018).
14. Harke, M. J. *et al.* Periodic and coordinated gene expression between a diazotroph and its diatom host. *ISME J.***13**, 118–131 (2019).
15. del Carmen Muñoz-Marin, M. *et al.* The transcriptional cycle is suited to daytime N2 fixation in the unicellular cyanobacterium 'Candidatus Atelocyanobacterium thalassa'(UCYN-A). *MBio***10**, (2019).
16. Vislova, A., Sosa, O. A., Eppley, J., Romano, A. & DeLong, E. F. Diel oscillation of microbial gene transcription declines with increasing depth in oligotrophic ocean waters. *Front. Microbiol.***10**, 2191 (2019).
17. Hernández Limón, M. D. *et al.* Transcriptional patterns of Emiliania huxleyi in the North Pacific Subtropical Gyre reveal the daily rhythms of its metabolic potential. *Environ. Microbiol.***22**, 381–396 (2020).
18. White, A. E., Barone, B., Letelier, R. M. & Karl, D. M. Productivity diagnosed from the diel cycle of particulate carbon in the North Pacific Subtropical Gyre. *Geophys. Res. Lett.***44**, 3752–3760 (2017).
19. Muratore, D. *et al.* Community-scale Synchronization and Temporal Partitioning of Gene Expression, Metabolism, and Lipid Biosynthesis in Oligotrophic Ocean Surface Waters. *bioRxiv* 2020.05.15.098020 (2020) doi:10.1101/2020.05.15.098020.
20. C-MORE : HOE-Legacy. https://hahana.soest.hawaii.edu/hoelegacy/hoelegacy.html.
21. Gifford, S. M., Becker, J. W., Sosa, O. A., Repeta, D. J. & DeLong, E. F. Quantitative Transcriptomics Reveals the Growth- and Nutrient-Dependent Response of a Streamlined Marine Methylotroph to Methanol and Naturally Occurring Dissolved Organic Matter. *MBio***7**, (2016).
22. Chevreux, B. *et al.* Using the miraEST assembler for reliable and automated mRNA transcript assembly and SNP detection in sequenced ESTs. *Genome Res.***14**, 1147–1159 (2004).
23. Roux, S., Enault, F., Hurwitz, B. L. & Sullivan, M. B. VirSorter: mining viral signal from microbial genomic data. *PeerJ***3**, e985 (2015).
24. Boetzer, M., Henkel, C. V., Jansen, H. J., Butler, D. & Pirovano, W. Scaffolding pre-assembled contigs using SSPACE. *Bioinformatics***27**, 578–579 (2011).
25. Krueger, F. *TrimGalore*. (Github).
26. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods***9**, 357–359 (2012).

27. Langfelder, P. & Horvath, S. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics***9**, 559 (2008).
28. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.***15**, 550 (2014).
29. Agrahari, R. *et al.* Applications of Bayesian network models in predicting types of hematological malignancies. *Sci. Rep.***8**, 6951 (2018).
30. Team, R. C. R: A language and environment for statistical com-puting. (2017).
31. Scutari, M., Scutari, M. M. & Mmpc, H.-P. Package 'bnlearn'. *Bayesian network structure learning, parameter learning and inference, R package version 4. 4***1**, (2019).
32. Shannon, P. *et al.* Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.***13**, 2498–2504 (2003).
33. Koller, D. & Friedman, N. *Probabilistic Graphical Models: Principles and Techniques*. (MIT Press, 2009).
34. Wozniak, D. J., Tiwari, K. B., Soufan, R. & Jayaswal, R. K. The mcsB gene of the clpC operon is required for stress tolerance and virulence in Staphylococcus aureus. *Microbiology***158**, 2568–2576 (2012).
35. Suzuki, C. K. *et al.* ATP-dependent proteases that also chaperone protein biogenesis. *Trends Biochem. Sci.***22**, 118–123 (1997).
36. Sowell, S. M. *et al.* Transport functions dominate the SAR11 metaproteome at low-nutrient extremes in the Sargasso Sea. *ISME J.***3**, 93–105 (2009).
37. Miller, D., Xu, H. & White, R. H. S-Inosyl-L-Homocysteine Hydrolase, a Novel Enzyme Involved in S-Adenosyl-L-Methionine Recycling. *J. Bacteriol.***197**, 2284–2291 (2015).
38. Smith, D. P. *et al.* Proteome Remodeling in Response to Sulfur Limitation in 'Candidatus Pelagibacter ubique'. *mSystems***1**, (2016).
39. Kiene, R. P., Hoffmann Williams, L. P. & Walker, J. E. Seawater microorganisms have a high affinity glycine betaine uptake system which also recognizes dimethylsulfoniopropionate. *Aquat. Microb. Ecol.***15**, 39–51 (1998).
40. Kiene, R. P. & Service, Susan K. Decomposition of dissolved DMSP and DMS in estuarine waters: dependence on temperature and substrate concentration. *Mar. Ecol. Prog. Ser.***76**, 1–11 (1991).
41. Díaz, A., Lacks, S. A. & López, P. The 5' to 3' exonuclease activity of DNA polymerase I is essential for Streptococcus pneumoniae. *Mol. Microbiol.***6**, 3009–3019 (1992).
42. Yoch, D. C., Ansede, J. H. & Rabinowitz, K. S. Evidence for Intracellular and Extracellular Dimethylsulfoniopropionate (DMSP) Lyases and DMSP Uptake Sites in Two Species of Marine Bacteria. *Appl. Environ. Microbiol.***63**, 4625 (1997).
43. Tripp, H. J. *et al.* SAR11 marine bacteria require exogenous reduced sulphur for growth. *Nature***452**, 741–744 (2008).

44. Dupont, C. L. *et al.* Genomic insights to SAR86, an abundant and uncultivated marine bacterial lineage. *ISME J.***6**, 1186–1199 (2012).
45. Wirth, J. S., Wang, T., Huang, Q., White, R. H. & Whitman, W. B. Dimethylsulfoniopropionate Sulfur and Methyl Carbon Assimilation in Ruegeria Species. *MBio***11**, (2020).
46. Tripp, H. J. The unique metabolism of SAR11 aquatic bacteria. *J. Microbiol.***51**, 147–153 (2013).
47. Steindler, L., Schwalbach, M. S., Smith, D. P., Chan, F. & Giovannoni, S. J. Energy starved Candidatus Pelagibacter ubique substitutes light-mediated ATP production for endogenous carbon respiration. *PLoS One***6**, e19725 (2011).
48. Olson, D. K., Yoshizawa, S., Boeuf, D., Iwasaki, W. & DeLong, E. F. Proteorhodopsin variability and distribution in the North Pacific Subtropical Gyre. *ISME J.***12**, 1047–1060 (2018).
49. Thompson, L. R. *et al.* Phage auxiliary metabolic genes and the redirection of cyanobacterial host carbon metabolism. *Proc. Natl. Acad. Sci. U. S. A.***108**, E757–64 (2011).
50. Wegrzyn, A., Czyz, A., Gabig, M. & Wegrzyn, G. ClpP/ClpX-mediated degradation of the bacteriophage lambda O protein and regulation of lambda phage and lambda plasmid replication. *Arch. Microbiol.***174**, 89–96 (2000).
51. Waldbauer, J. R. *et al.* Nitrogen sourcing during viral infection of marine cyanobacteria. *Proc. Natl. Acad. Sci. U. S. A.***116**, 15590–15595 (2019).
52. Cuhel, R. L., Ortner, P. B. & Lean, D. R. S. Night synthesis of protein by algae 1. *Limnol. Oceanogr.***29**, 731–744 (1984).
53. Carini, P., Steindler, L., Beszteri, S. & Giovannoni, S. J. Nutrient requirements for growth of the extreme oligotroph 'Candidatus Pelagibacter ubique'HTCC1062 on a defined medium. *ISME J.***7**, 592–602 (2013).
54. Giovannoni, S. J. *et al.* Genome streamlining in a cosmopolitan oceanic bacterium. *Science***309**, 1242–1245 (2005).
55. Aylward, F. O. *et al.* Microbial community transcriptional networks are conserved in three domains at ocean basin scales. *Proc. Natl. Acad. Sci. U. S. A.***112**, 5443–5448 (2015).

Figures

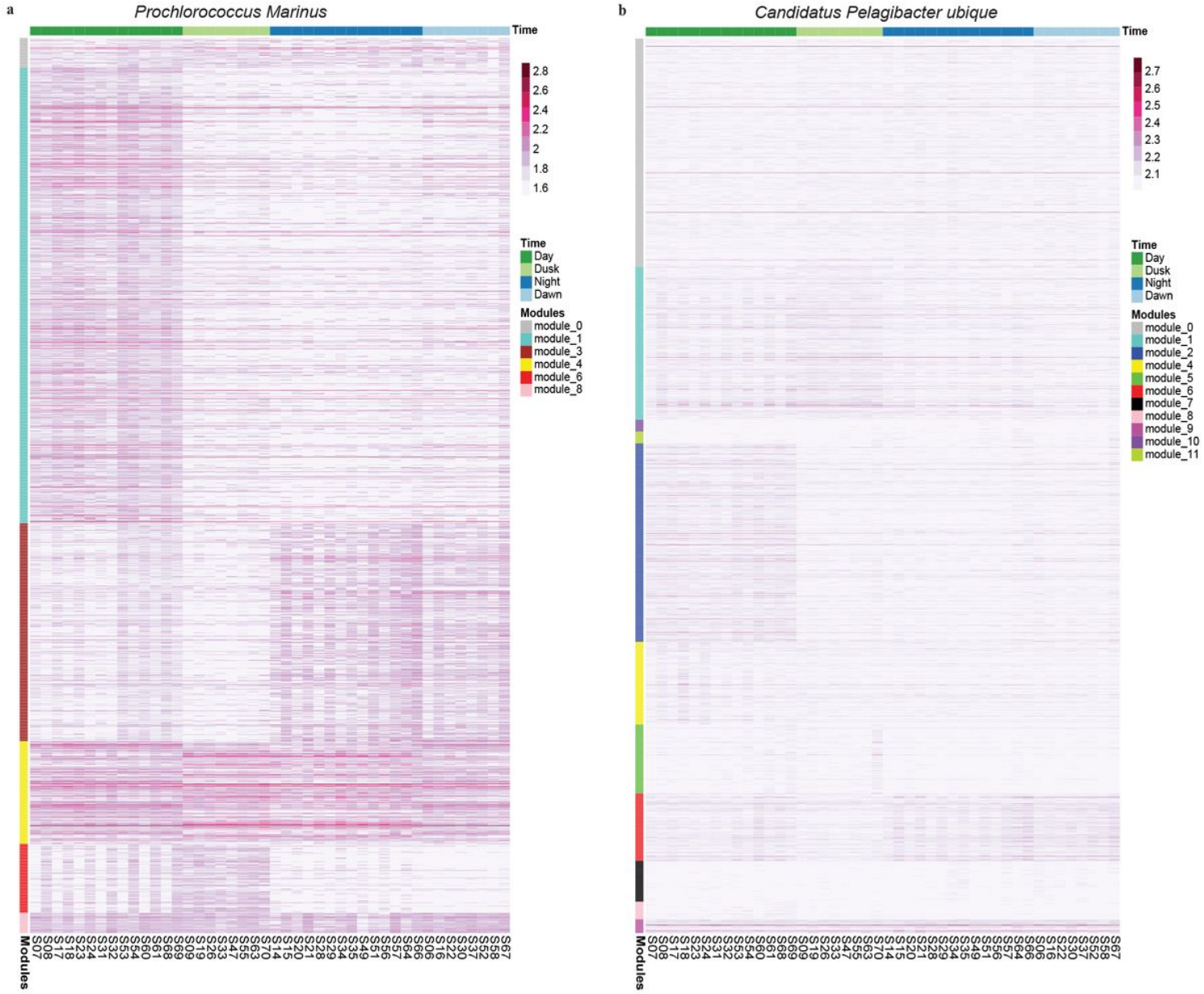

Figure 1

Gene module expression profile of the two species. a: expression profile of Prochlorococcus marinus; b: expression profile of Candidatus Pelagibacter ubique. Each column corresponds to a sample, and each row represents different genes. Gene modules are clustered based on the similarity of expression. Samples are grouped by sampling time of the day (Day: 10:00 and 14:00; Dusk: 18:00; Night: 22:00 and 2:00; Dawn: 6:00). Different sample groups are shown at the top of the heatmap.

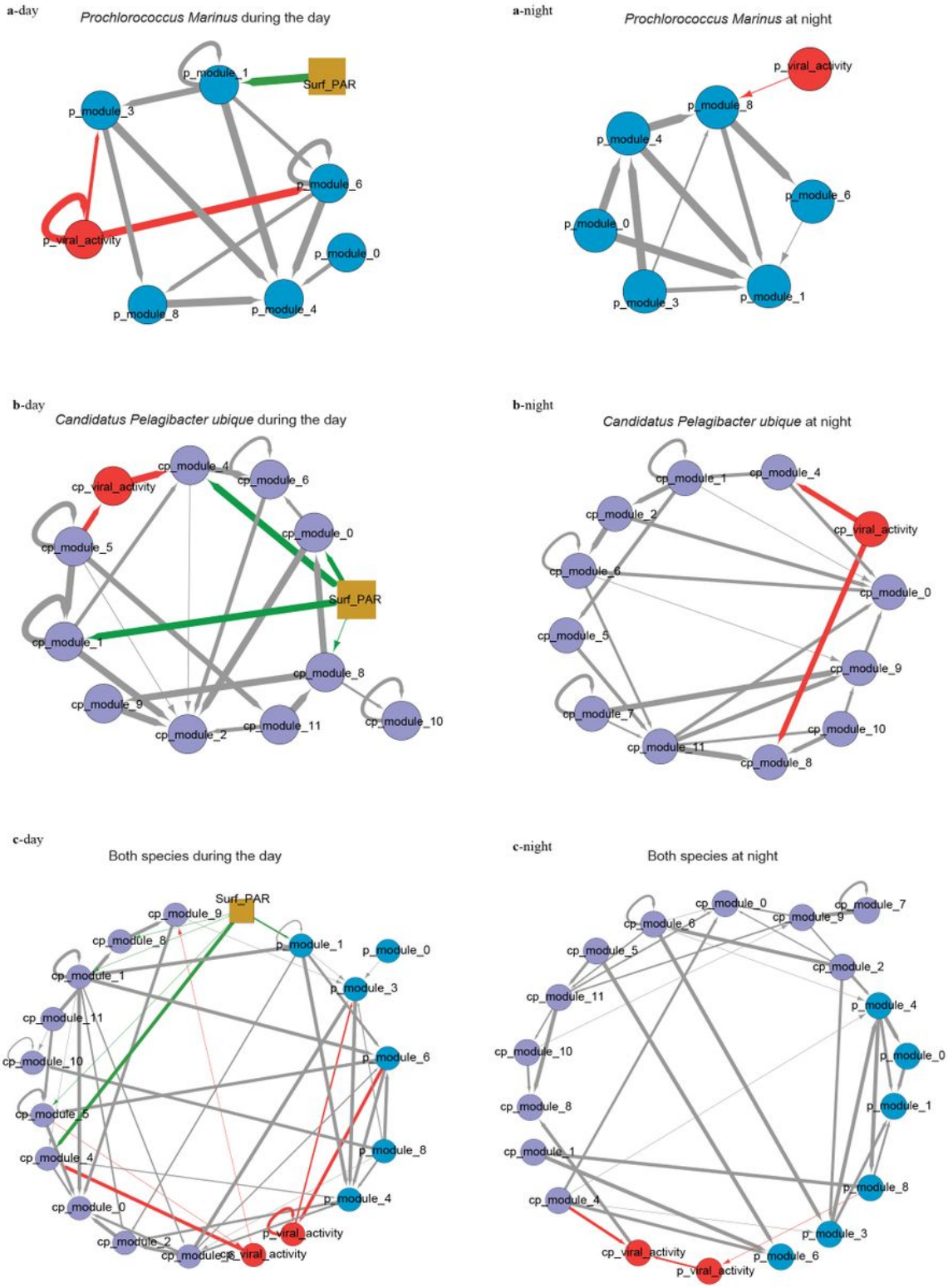

Figure 2

Averaged Dynamic Bayesian Network structures. a-day: Averaged DBN structure of Prochlorococcus marinus during the day. a-night: Averaged DBN structure of Prochlorococcus marinus at night. b-day: Averaged DBN structure of Candidatus Pelagibacter ubique during the day. b-night: Averaged DBN structure of Candidatus Pelagibacter ubique during the night. c-day: Averaged DBN structure of both species combined during the day. c-night: Averaged DBN structure of both species combined during the

night. Blue circle nodes are Prochlorococcus marinus gene modules; purple circle nodes are Candidatus Pelagibacter ubique gene modules; red circle nodes are viral activities; and brown square nodes are surface PAR light intensity. Nodes with edges connected to themselves indicated they are affected by their previous time point. Red edges indicate the connection between viral activity and other nodes. Green edges indicate connection between environmental variables and other nodes.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- Supplementarydatasetsnotes.docx
- Supplementarydataset1.xlsx
- Supplementarydataset2.xlsx
- Supplementarydataset3.xlsx
- Supplementarydataset4.xlsx
- Supplementarydataset5.xlsx
- SupplementaryFigures.pdf
- SupplementaryTables.pdf