# *Multispecies diel transcriptional oscillations in open ocean heterotrophic bacterial assemblages*



**Citation:** Ottesen, E. A., C. R. Young, S. M. Gifford, J. M. Eppley, R. Marin, S. C. Schuster, C. A. Scholin, and E. F. DeLong. "Multispecies Diel Transcriptional Oscillations in Open Ocean Heterotrophic Bacterial Assemblages." Science 345, no. 6193 (July 11, 2014): 207–212.

**As Published:** http://dx.doi.org/10.1126/science.1252476

**Publisher:** American Association for the Advancement of Science (AAAS)

**Persistent URL:** http://hdl.handle.net/1721.1/88413

**Version:** Author's final manuscript: final author's manuscript post peer review, without publisher's formatting or copy editing

# Title: Multispecies diel transcriptional oscillations in open ocean heterotrophic bacterial assemblages

**Authors:** Elizabeth A. Ottesen[1,2,3], Curtis R. Young[1,2], Scott M. Gifford[1,2], John M. Eppley[1,2], Roman Marin III[4], Stephan C. Schuster[5], Christopher A. Scholin[4] and Edward F. DeLong[1,2,6].

CONFIDENTIAL: Accepted, Science magazine, 5/22/2014:

**Affiliations:**

[1]Department of Civil and Environmental Engineering, Massachusetts Institute of Technology, Cambridge, MA 02139, USA.

[2]Center for Microbial Oceanography: Research and Education, University of Hawaii, Honolulu, HI 96822, USA.

[3]Department of Microbiology, University of Georgia**,** Athens, GA 30602, USA.

[4]Monterey Bay Aquarium Research Institute, Moss Landing, CA 95039, USA.

[5]Singapore Centre on Environmental Life Sciences Engineering, Nanyang Technological University, Singapore 637551.

[6] Department of Biological Engineering, Massachusetts Institute of Technology, Cambridge, MA 02139, USA.

*Correspondence to: delong@mit.edu; edelong@hawaii.edu

**Abstract**: Oscillating diurnal rhythms of gene transcription, metabolic activity and behavior are found in all three domains of life. Diel cycles in naturally occurring heterotrophic bacteria and archaea however, have rarely been observed. Here we report time-resolved whole genome transcriptome profiles of multiple, naturally occurring oceanic bacterial populations sampled *in situ* over three days. As anticipated, the cyanobacterial transcriptome exhibited pronounced diel periodicity. Unexpectedly however, several different heterotrophic bacterioplankton groups also displayed diel cycling in many of their gene transcripts. Furthermore, diel oscillations in different heterotrophic bacterial groups suggested population-specific timing of peak transcript expression in a variety of metabolic gene suites. These staggered multispecies waves of diel gene transcription may influence both the tempo and mode of matter and energy transformation in the sea.

**Main Text:**

The coordination of biological activities into daily periodic cycles is a common feature of eukaryotes and is widespread among plants, fungi, and animals, including man (*1*). Among single celled non-eukaryotic microbes, diel cycles have been well documented in cyanobacterial isolates (*2-4*), one halophilic archaeon (*5*), and in bacterial symbionts of fish and squid (*6, 7*). Some evidence for diel cycling in microbial plankton has also been suggested on the basis of bulk community amino acid incorporation, viral production, or metabolite consumption (*8-10*). The existence of regular diel oscillations in free-living heterotrophic bacterial species however, has rarely been assessed.

Microbial community RNA sequencing techniques now allow simultaneous determination of whole genome transcriptome profiles among multiple co-occurring species (*11, 12*), enabling high frequency, time resolved analyses of microbial community dynamics (*12, 13*). To better understand temporal transcriptional dynamics in oligotrophic bacterioplankton communities, we conducted a high-resolution multi-day time series of bacterioplankton sampled from the North Pacific Subtropical Gyre (*14*).

To facilitate repeated sampling of the same planktonic microbial populations through time, automated Lagrangian sampling of bacterioplankton was performed every two hours over three days using a free-drifting robotic Environmental Sample Processor (ESP; (*13, 15*); Fig. S1). Following instrument recovery, planktonic microbial RNA was extracted, purified and converted to cDNA to assess whole genome transcriptome dynamics of predominant planktonic microbial populations (Table S1, Table S2). The recovered cDNAs were dominated by transcripts from *Prochlorococcus* and several proteorhodopsin-containing or photoheterotrophic bacteria,

including members of the *Pelagibacter* (SAR11), *Roseobacter,* SAR116, SAR86, and SAR324 clades (Fig. S2).

Phylogenetic analysis of gene transcripts in the most abundant taxa revealed the presence of some microdiversity (Figs. S3-8). The most abundant transcripts sampled at any given time point however, were dominated by only a few genotypes within each population that persisted throughout the sampling period. An exception was *Roseobacter*, with transcripts for two different genes (*groEL* and *dnaK*) indicating the presence of a genotype that started at a very low abundance and increased in representation over the course of the time series. This variability could be due to an injection of a new population as water masses mixed during the latter portion of the time series, or possibly to an alteration in the relative transcriptional activities of two ecotypes that are responding to changes in the surrounding environment.

Transcriptional activity in *Prochlorococcus* was highly dependent on the time of day. Harmonic regression analyses indicated that nearly half (1,491) of all *Procholorococcus* population transcripts were significantly periodic (Table 1; Table S3; Fig 1). The expression patterns observed were similar to those of monocultures growing in controlled laboratory settings (*4*) but there were also notable differences (Fig. 1). For example, photosystem I gene expression exhibited a double peak in the wild *Prochlorococcus* transcriptome around noon (Fig. 1). In contrast, under laboratory conditions most photosystem I genes, i.e. *psaL* and *psaF*, were found to peak just before noon, while *psaA* and *psaB* peaked shortly after noon (*4*).

The largest discrepancy between *Prochlorococcus* laboratory studies and our field observations was that a considerable number of *Prochlorococcus* transcripts in our field populations peaked around midday (Fig. 1). Some of these genes did exhibit periodicity in cultures, but peaked at a different time of day in field populations. A larger fraction of these

mid-day peaking transcripts were either not periodically expressed, or were not present in the culture experiments. In addition, 62% of the 10am - 4pm peaking transcripts in our field study lacked KEGG orthology annotations, as opposed to those peaking in the evening or late at night.

A number of factors may be responsible for differences in transcript dynamics between in laboratory cultures versus field *Prochlorococcus* populations. Maximal light levels at our study site at 23 m depth were frequently two fold higher (450 umol $Q/m^{-2}/s^{-1}$) than those used in laboratory microarray experiments (232 umol $Q/m^{-2}/s^{-1}$) (*4*). Fundamental genetic differences between our field populations and the *Prochlorococcus* strain used in laboratory culture experiments likely also contribute to the differences we observed. Other variables, including nutrient composition and organismal interactions, may also be a factor in the observed differences. While we could not identify obvious trends in the type or function of transcripts showing peak expression during the mid-day period, they did include a wide range of enzymatic functions that are more consistent with nutrient-responsive metabolic changes rather than a simple high-light stress response.

An abundant *Roseobacter* population also showed strong diel oscillations in its transcriptome profile, most notably in expressed genes involved in bacteriochlorophyll-associated aerobic anoxygenic photosynthesis (AAnP). Overall, a large fraction of *Roseobacter* transcripts were periodically expressed (Table 1). Of these, the majority peaked during daylight hours, with only a few gene transcripts peaking at night (Fig. 2). While this pattern contrasts with that observed in *Prochlorococcus*, where most diurnally regulated transcripts peaked at dawn or dusk, it was consistent with transcriptional regulation recently reported in *Dinoroseobacter shibae* (*16*).

Thirty five of the forty significantly periodic *Roseobacter* transcripts that peaked between 11 pm and 7 am encoded genes belonging to a large photosynthetic "superoperon" (Fig. S9). Nightly expression of these genes, followed by immediate repression upon light onset, is consistent with the *D. shibae* study (*16*), and may be preparing cells for efficient solar energy harvest in the early morning hours. Functions that peaked during the day-time hours included ribosomal proteins, respiratory transcripts, genes involved in amino acid metabolism, and transporters (Fig. 2).

Proteorhodopsin-containing photoheterotrophs including members of the SAR11, SAR116, and SAR86 also showed evidence of diel periodicity in many of their gene transcripts (Fig. 2, Table 1). Interestingly, all opsin-containing bacteria analyzed (SAR11, SAR116, SAR86, and SAR324) exhibited statistically significant diel oscillations in their proteorhodopsin gene transcripts (Table S3; Fig. S10). Peak expression of the opsin transcripts occurred near dawn in all these populations (Fig. S10), potentially optimizing solar energy capture by the light-driven, proton-pumping rhodopsins.

Principal components analysis distinguished time series samples for each heterotroph by time of day (Fig. 3) and showed significant correlation with the light-driven behavior of *Prochlorococcus* (Table 1). Overall, this data is consistent with profound, genome-wide transcriptional changes across the day-night cycle for each population. In addition, co-clustering of transcripts using GeneARMA (GA; (*14, 17*)) revealed suites of gene transcripts that exhibited similar expression patterns among different taxa (Fig. 4; Fig. S11-14; Table S4). For example, a group of transcripts that that fit highly similar GA expression models across multiple species included Pro GA5, Pro GA7, Pro GA9, Pro GA23, SAR11 GA6. SAR11 GA18, SAR116 GA2, and *Roseobacter* GA8 (Fig 4; Fig S14; Table S4). These multispecies, day-peaking transcripts

(Fig S14; Table S4) included gene products associated with respiration (*Prochlorococcus*, SAR11, SAR116, *Roseobacter*), nitrogen metabolism (*Prochlorococcus*, SAR11, SAR116), glycine metabolism (*Prochlorococcus*, SAR11, *Roseobacter*), carbon monoxide metabolism (SAR116, *Roseobacter*) and DNA synthesis (*Prochlorococcus*, *Roseobacter*). This co-clustering of gene transcripts reveals a complex pattern of expression through the day and across the time series, and provides evidence for parallel trends in gene expression across multiple species (Fig. 4, Fig. S11-14, Table S4).

Together, the transcriptional profiles of *Roseobacter*, SAR11, SAR116 and SAR86 indicate diel cycling of metabolic gene transcripts, and suggest a multispecies wave-like progression of upregulated gene suites across the day/night cycle (Fig. 4). Most conspicuously, a regular diel succession of translational, transcriptional and respiratory gene transcripts was followed by peaks in transporter transcripts that possibly reflect a metabolic recovery phase (Fig. 2). Many of these metabolic pathway transcripts peaked earlier in the day in *Roseobacter* field populations relative to other bacterial heterotrophs (Fig. 2, Fig 4, Table S4).

The overall transcriptional profile of SAR324 did not show as many transcript diel oscillations as other heterotrophic taxa. Instead, principal components analysis clustered SAR324 transcripts according to the day that they were collected (Fig. 3). In particular, the SAR324 group showed a strong separation between the first portion of the time series and the second in principal components analysis (Fig. 3). This split appears to be associated with the increases in temperature and salinity observed across the time series (Fig. S1).

The diurnal patterns reported here for open ocean heterotrophic bacterioplankton were different from those observed in a previous study of phylogenetically related coastal bacterioplankton using similar methods (*12*). For example, coastal versus open ocean SAR11

147 populations revealed differential expression levels among several orthologous transcript
148 categories (Fig. S15). Additionally, while the open ocean SAR11 populations reported here
149 exhibited statistically significant diel oscillations for many gene transcripts (Fig. 2), the coastal
150 SAR11 populations did not.

151 Currently available data are insufficient to provide definitive mechanistic explanations
152 for the diel behaviors we observed in different heterotrophic bacterioplankton species. It is
153 possible that photoreceptors in these bacteria are involved in regulating light-dark cycles of
154 transcriptional activity. Marine *Roseobacter* species have previously been shown to regulate
155 their global transcriptional behavior in response to light (*16*), and laboratory cultures of
156 *Pelagibacter* also exhibit light-responsive metabolic behaviors (*18*). Differences between the
157 behaviors of SAR11 coastal versus open ocean field populations however (Fig. S15), as well as
158 comparisons of several taxa in our field study versus laboratory experiments on related cultivated
159 isolates (Fig. 1), suggest that other factors may be at play in regulating diel behavior among these
160 different bacterioplankton populations.

161 Previous studies have proposed that tight metabolic coupling between primary producers
162 and consumers in microbial plankton might elicit conspicuous diel cycling in heterotrophic
163 bacterial activities (*8*). The diel cycling we observed among different bacterioplankton species is
164 consistent with this hypothesis, with multiple co-existing heterotroph populations exhibiting
165 diurnal oscillations resembling those of their photoautotrophic neighbors. We postulate that the
166 tightly coupled multispecies temporal expression patterns observed may elicit corresponding
167 waves of species-specific metabolic responses at regular time intervals, potentially coordinating
168 diverse biogeochemical activities in these complex microbial communities. Such temporal
169 coordination of biogeochemical activities among multiple species may be important regulators of

both the tempo and mode of microbial matter and energy transformation in the sea.

## References and Notes:


1. D. Bell-Pedersen *et al.*, Circadian rhythms from multiple oscillators: lessons from diverse organisms. *Nature reviews. Genetics* **6**, 544 (2005).
2. I. M. Axmann, S. Hertel, A. Wiegard, A. K. Dorrich, A. Wilde, Diversity of KaiC-based timing systems in marine Cyanobacteria. *Marine genomics* **In press,** http://dx.doi.org/10.1016/j.margen.2013.12.006, (2014).
3. H. Ito *et al.*, Cyanobacterial daily life with Kai-based circadian and diurnal genome-wide transcriptional control in *Synechococcus elongatus*. *Proc. Natl. Acad. Sci. U.S.A.* **106**, 14168 (2009).
4. E. R. Zinser *et al.*, Choreography of the transcriptome, photophysiology, and cell cycle of a minimal photoautotroph, prochlorococcus. *PloS one* **4**, e5135 (2009).
5. K. Whitehead, M. Pan, K. Masumura, R. Bonneau, N. S. Baliga, Diurnally entrained anticipatory behavior in archaea. *PloS one* **4**, e5485 (2009).
6. J. F. Flint, D. Drzymalski, W. L. Montgomery, G. Southam, E. R. Angert, Nocturnal production of endospores in natural populations of *Epulopiscium*-like surgeonfish symbionts. *Journal of bacteriology* **187**, 7460 (2005).
7. A. M. Wier *et al.*, Transcriptional patterns in both host and bacterium underlie a daily rhythm of anatomical and metabolic change in a beneficial symbiosis. *Proc. Natl. Acad. Sci. U.S.A.* **107**, 2259 (2010).
8. J. M. Gasol *et al.*, Diel variations in bacterial heterotrophic activity and growth in the northwestern Mediterranean Sea. *Mar Ecol Prog Ser* **164**, 107 (1998).
9. C. Winter, G. J. Herndl, M. G. Weinbauer, Diel cycles in viral infection of bacterioplankton in the North Sea. *Aquat Microb Ecol* **35**, 207 (2004).
10. M. Galí *et al.*, Diel patterns of oceanic dimethylsulfide (DMS) cycling: Microbial and physical drivers. *Global Biogeochem Cycles* **27**, 620 (2013).
11. S. M. Gifford, S. Sharma, M. Booth, M. A. Moran, Expression patterns reveal niche diversification in a marine microbial assemblage. *The ISME journal* **7**, 281 (2013).
12. E. A. Ottesen *et al.*, Pattern and synchrony of gene expression among sympatric marine microbial populations. *Proc. Natl. Acad. Sci. U.S.A.* **110**, E488 (2013).
13. E. A. Ottesen *et al.*, Metatranscriptomic analysis of autonomously collected and preserved marine bacterioplankton. *The ISME journal* **5**, 1881 (2011).
14. Materials and methods are available as supplementary material on Science Online
15. C. M. Preston *et al.*, Underwater application of quantitative PCR on an ocean mooring. *PloS one* **6**, e22522 (2011).
16. J. Tomasch, R. Gohl, B. Bunk, M. S. Diez, I. Wagner-Döbler, Transcriptional response of the photoheterotrophic marine bacterium *Dinoroseobacter shibae* to changing light regimes. *The ISME journal* **5**, 1957 (2011).
17. N. Li *et al.*, Functional clustering of periodic transcriptional profiles through ARMA(*p,q*). *PloS one* **5**, e9894 (2010).
18. L. Steindler, M. S. Schwalbach, D. P. Smith, F. Chan, S. J. Giovannoni, Energy starved Candidatus Pelagibacter ubique substitutes light-mediated ATP production for endogenous carbon respiration. *PloS one* **6**, e19725 (2011).
19. Z. Liu, W. Hsiao, B. L. Cantarel, E. F. Drábek, C. Fraser-Liggett, Sparse distance-based learning for simultaneous multiclass classification and feature selection of metagenomic data. *Bioinformatics* **27**, 3242 (2011).
20. J. Okansen et al. Package "vegan": Community ecology package. http://cran.rproject.org/web/packages/vegan//vegan.pdf (2012).
21. J. C. Robidart *et al.*, Ecogenomic sensor reveals controls on N-fixing microorganisms in the North Pacific Ocean. *The ISME journal* doi: 10.1038/ismej.2013.244, (2014).
22. F. J. Stewart, E. A. Ottesen, E. F. DeLong, Development and quantitative analyses of a universal rRNA-subtraction protocol for microbial metatranscriptomics. *The ISME journal* **4**, 896 (2010).

23. Y. Shi, G. W. Tyson, J. M. Eppley, E. F. DeLong, Integrated metatranscriptomic and metagenomic analyses of stratified microbial assemblages in the open ocean. *The ISME journal* **5**, 999 (2011).
24. S. Rodrigue *et al.*, Unlocking short read sequencing for metagenomics. *PloS one* **5**, e11840 (2010).
25. S. M. Kielbasa, R. Wan, K. Sato, P. Horton, M. C. Frith, Adaptive seeds tame genomic sequence comparison. *Genome research* **21**, 487 (2011).
26. B. K. Swan *et al.*, Prevalent genome streamlining and latitudinal divergence of planktonic bacteria in the surface ocean. *Proc. Natl. Acad. Sci. U.S.A.* **110**, 11463 (2013).
27. R. J. Newton *et al.*, Genome characteristics of a generalist marine bacterial lineage. *The ISME journal* **4**, 784 (2010).
28. M. Kanehisa, S. Goto, KEGG: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* **28**, 27 (2000).
29. Team RDC (2012) R: A Language and Environment for Statistical Computing (R Foundation for Statistical Computing, Vienna).
30. Y. Benjamini, Y. Hochberg, Controlling the false discovery rate - a practical and powerful approach to multiple testing. *J Roy Stat Soc B Met* **57**, 289 (1995).
31. J. P. Baudry, A. E. Raftery, G. Celeux, K. Lo, R. Gottardo, Combining Mixture Components for Clustering. *Journal of computational and graphical statistics : a joint publication of American Statistical Association, Institute of Mathematical Statistics, Interface Foundation of North America* **9**, 332 (2010).
32. R. C. Edgar, Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* **26**, 2460 (2010).
33. R. C. Edgar, MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **32**, 1792 (2004).
34. M. N. Price, P. S. Dehal, A. P. Arkin, FastTree: computing large minimum evolution trees with profiles instead of a distance matrix. *Molecular biology and evolution* **26**, 1641 (2009).
35. J. G. Caporaso *et al.*, QIIME allows analysis of high-throughput community sequencing data. *Nature methods* **7**, 335 (2010).

## Acknowledgments:

We thank the officers and crew of the Kilo Moana, Chief Scientist Julie Robidart, Sam Wilson, and the ESP engineering and Science team (John Ryan, Jim Birch, Chris Preston, Gene Massion, Scott Jensen, Brent Roman) for all the able assistance. This work was supported by grants from the Gordon and Betty Moore Foundation GBMF #492.01 and #3777 (EFD) and NSF grant EF0424599 (EFD). Development of the ESP was supported by National Science Foundation Grant OCE-0314222 (to C.A.S.), National Aeronautics and Space Administration Astrobiology Grants NNG06GB34G and NNX09AB78G (to C.A.S.), the Gordon and Betty Moore Foundation (C.A.S.), and the David and Lucile Packard Foundation. This work is a contribution of the Center for Microbial Oceanography: Research and Education (C-MORE). Sequences reported in

this paper have been deposited in the Gen-Bank database (accession no. SRP041215).

**Supplementary Materials:**

Materials and Methods

Figures S1-S14

Table S1

Legends for External Tables S2-S4

References (*27-43*)

**Fig. 1.** Laboratory versus field comparisons of periodic expression patterns in *Prochlorococcus* populations. **A.** Scatter plot shows time of peak abundance for 973 transcripts identified as significantly periodic in both studies. Histograms show the total number of genes peaking in 1-hour intervals in this study (top) and the laboratory experiment (side). Black bars represent genes identified as significantly periodic in both studies, grey bars represent genes expressed in both studies but significantly periodic in only one, and white bars represent significantly periodic transcripts that were not detected in the other dataset. For this comparison, we used published significance cutoffs from the laboratory study (*4*), but for consistency generated new peak times using our harmonic regression approach and the published normalized mean expression levels for each time point. In general, the peak times generated using our approach closely matched published values for that dataset. **B-C.** Plots showing relative expression (normalized to mean expression level) over time for our metatranscriptome (top trace) and in microarray data (bottom trace) for selected transcripts. For comparison, experimental midnights (24 hr and 48hr) from the microarray study are aligned with the 12:00AM samples from 9/9 and 9/10, respectively. All ATP synthase subunits (**B**) and selected subunits from photosystem I (**C**) are shown.

**Fig. 2. Timing of periodically expressed transcripts.** For each population, a histogram showing the number of periodically expressed diel transcripts with peak expression within 1-hr intervals throughout the day is shown (left). On the right, time of peak expression of all transcripts assigned to selected KEGG pathways is plotted (grey). Red circles denote transcripts identified as significantly periodic (24 hour period). The "transporters" category includes both the ABC Transporters KEGG pathway and the Transporters BRITE hierarchy, the "photosynthesis" category includes both the Photosynthesis KEGG pathway and the BRITE Photosynthesis Proteins categorizations. "Carbon Fixation" refers to genes assigned to the Carbon Fixation in Photosynthetic Organisms KEGG Pathway. The photosynthesis and carbon fixation categories are present in heterotrophic organisms due to cross-assignment of ATP Synthase genes and pentose phosphate cycle genes. Black and yellow bars depict the daily photoperiod (based on sunrise and sunset times).

**Fig. 3.** Principal Component analysis of population transcriptional profiles. Transcript abundances were normalized to total transcripts assigned to each population at each time point,

and arcsin transformed to approximate normality (*19*). Symbol color denotes time of day and shape denotes day of collection. Grey lines connect samples to centroids for selected sample groupings that separate points well. *Roseobacter* SAGs: samples collected between 7am and 9pm (vs. 9pm to 7am); SAR11, SAR116 and SAR86 cluster, samples collected between 9am and 6pm (and vice versa); SAR324 cluster, samples collected before or after 9/9 4pm. All factor correlations shown were highly significant ($p = 0.001$). Alternative time of day categories were also highly significant for Roseobacter SAGs, SAR11, SAR116 and SAR86. SAR116 ($r^2$ 0.10, p 0.037) and SAR86 ($r^2$ 0.12, p 0.019) also correlated weakly with the grouping shown for SAR324. All analyses carried out using functions in the vegan software package (*20*).

**Fig. 4**. Timing of expression of functional gene clusters in different taxa clustered by similarity. Heatmap shows cluster models for all geneARMA clusters, colored by mean-centered relative expression (red=high, blue=low). Black and yellow bars show the daily photoperiod. Each box represents a single sampling event, for sample times see Table S1. Dendrograms show cluster model similarity (Pearson correlations, average linkage clustering, scale bar at upper right represents a correlation of 0.5). The total number of genes (A), significantly periodic genes (B), and genes associated with Photosynthesis (C), Ribosome (D), Oxidative Phosphoryloation (E), Amino Acid Metabolism (F), and Transport (G) (defined as for Figure 2), are listed for each cluster.

Table 1: Harmonic Regression Results

| | Prochlorococcus | Roseobacter SAGs | SAR11 | SAR116 | SAR86 | SAR324 |
|---|---|---|---|---|---|---|
| Sequence reads[1] | 2886677 | 177982 | 774064 | 200368 | 151468 | 118098 |
| Transcripts[2] | 3045 | 2604 | 2802 | 2618 | 2367 | 4732 |
| Periodic[3] | 1491 | 426 | 201 | 80 | 10 | 8 |
| Constrained PCA vs 24-hour clock[4] | 0.68 (p = 0.005) | 0.49 (p = 0.005) | 0.24 (p = 0.005) | 0.15 (p = 0.005) | 0.13 (p = 0.005) | 0.10 (p = 0.01) |
| Procrustes Test vs. Prochlorococcus PCA[5] | | 0.78 (p < 0.001) | 0.55 (p < 0.001) | 0.70 (p < 0.001) | 0.52 (p < 0.001) | 0.36 (p = 0.031) |
| Mantel Test vs. Prochlorococcus[6] | | 0.63 (p < 0.001) | 0.40 (p < 0.001) | 0.31 (p = 0.003) | 0.26 (p = 0.005) | 0.27 (p = 0.002) |

[1] The total number of sequence reads assigned to each taxon bin

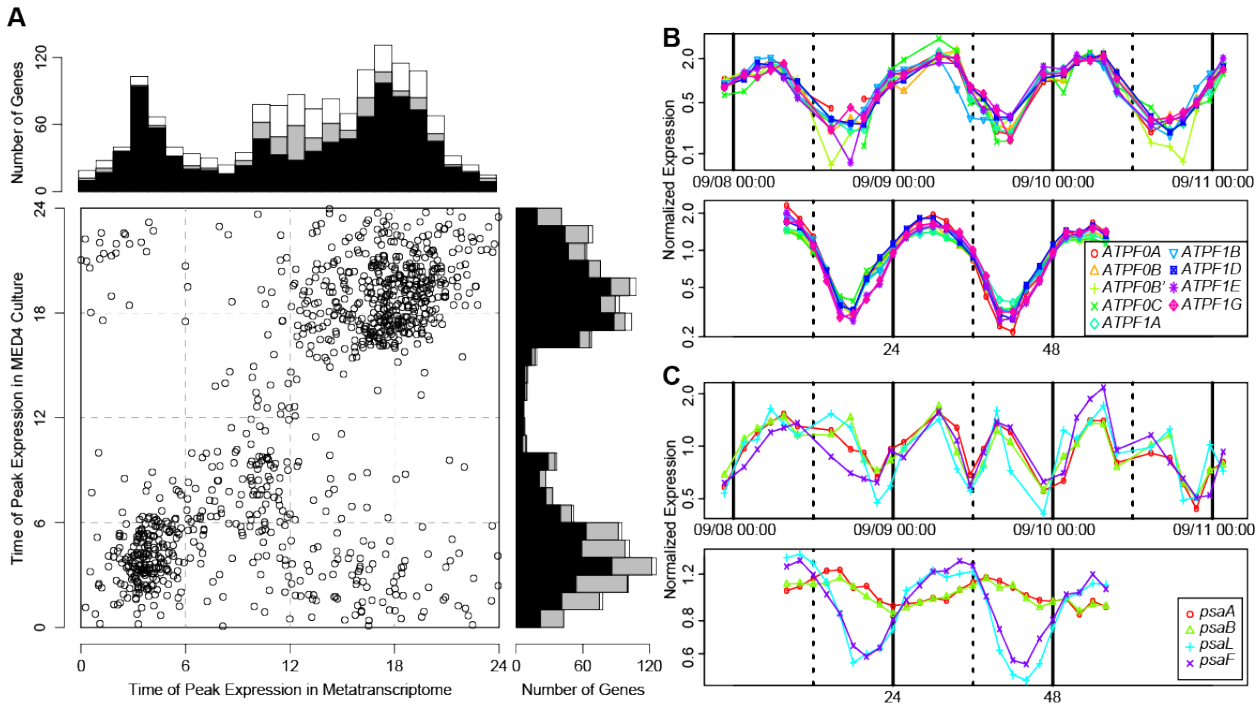
[2] The total number of unique ortholog clusters (see Database S1) with at least one mapped sequence.

[3] The total number of sequences identified as showing 24-hour periodicity using harmonic regression.

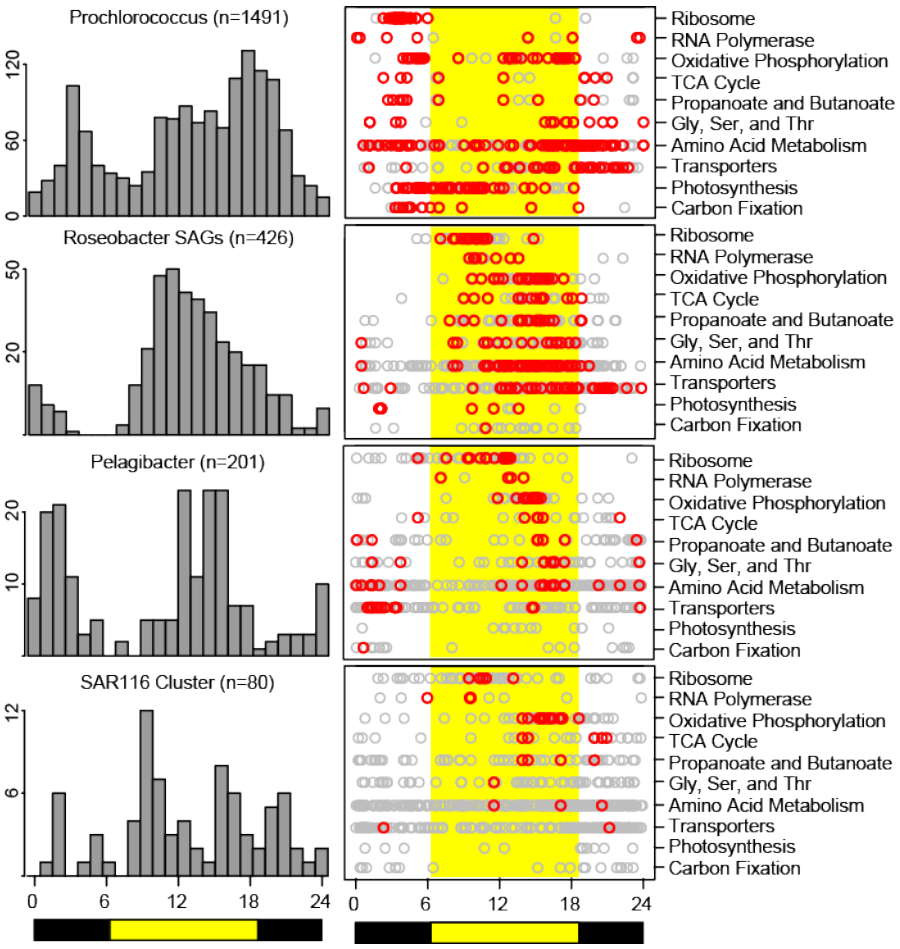
[4] Proportion of variance explained by 24-hour periodicity in constrained principal components analysis.

[5] Procrustes correlation between the first two principal components from *Prochlorococcus* and other taxa (unconstrained principal components analysis as shown in Fig. 3). P-value based on 999 permutations.

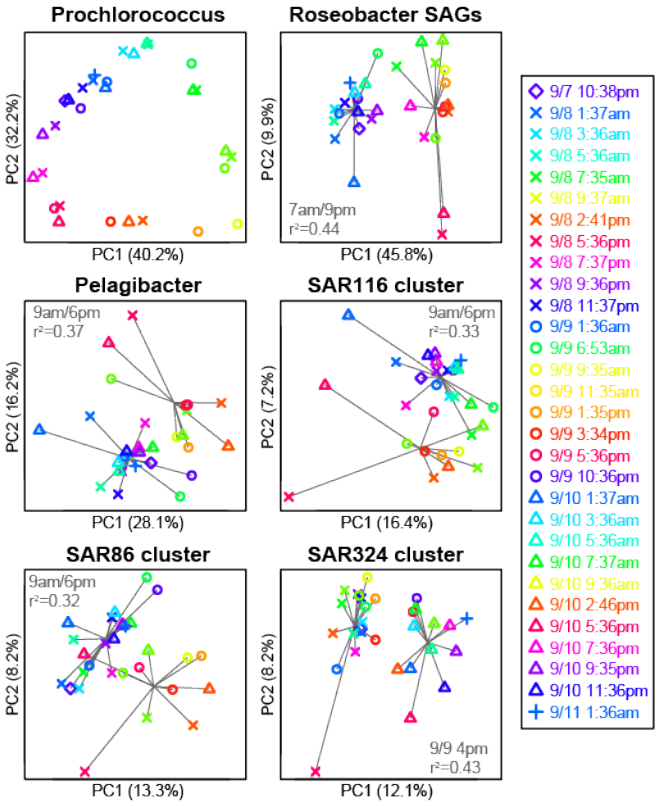
[6] Correlation between pairwise sample similarities from *Prochlorococcus* and heterotrophic taxa based on Mantel test on Euclidean distance matrices. P-value based on 999 permutations.


A
Number of Genes
Time of Peak Expression in MED4 Culture
Time of Peak Expression in Metatranscriptome
Number of Genes
B
Normalized Expression
09/08 00:00
09/09 00:00
09/10 00:00
09/11 00:00
ATPF0A
ATPF0B
ATPF0B'
ATPF0C
ATPF1A
ATPF1B
ATPF1D
ATPF1E
ATPF1G
C
psaA
psaB
psaL
psaF

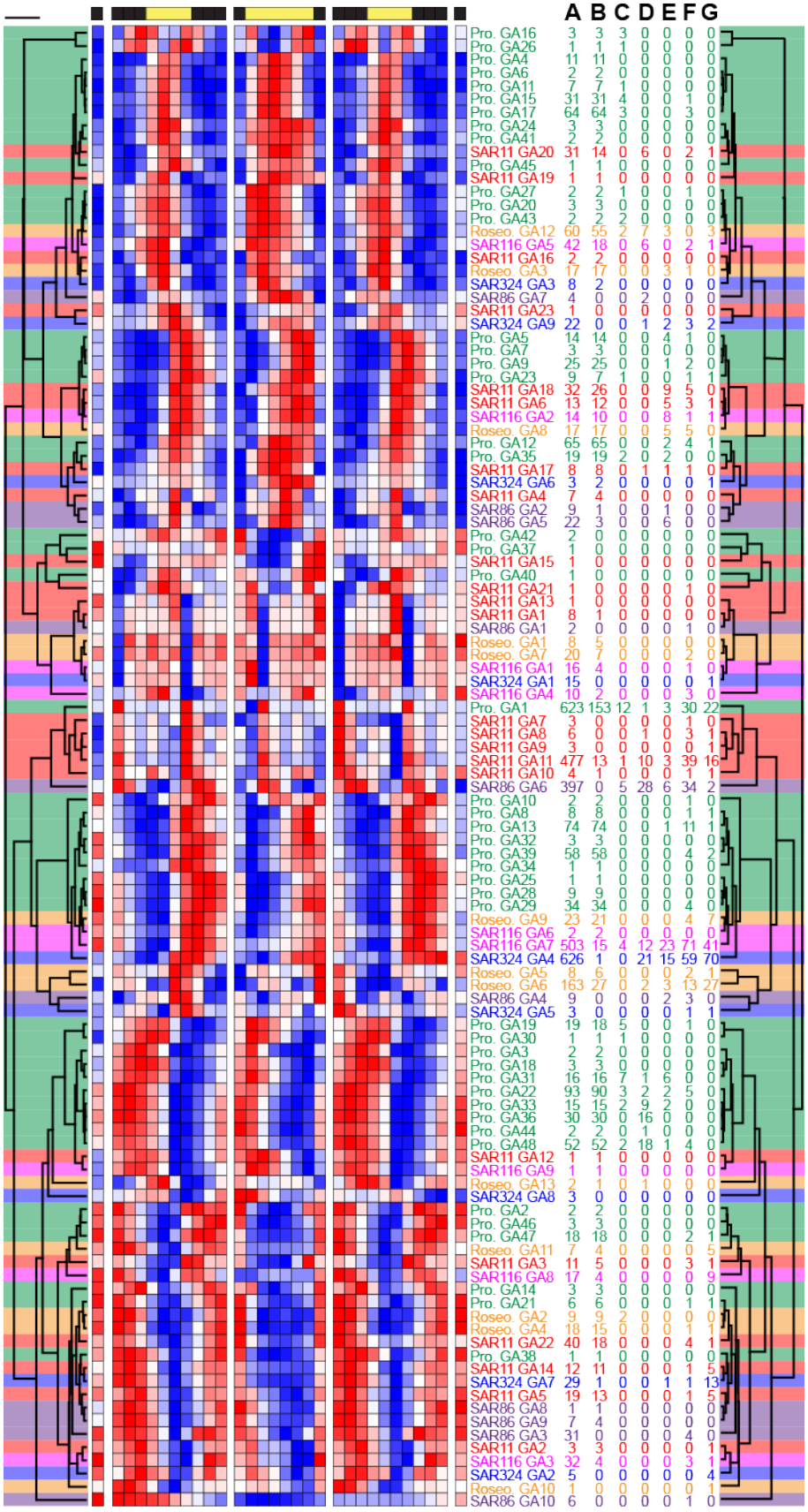
332


Prochlorococcus (n=1491)
Roseobacter SAGs (n=426)
Pelagibacter (n=201)
SAR116 Cluster (n=80)
Ribosome
RNA Polymerase
Oxidative Phosphorylation
TCA Cycle
Propanoate and Butanoate
Gly, Ser, and Thr
Amino Acid Metabolism
Transporters
Photosynthesis
Carbon Fixation
0
6
12
18
24

333
334


Prochlorococcus
PC2 (32.2%)
PC1 (40.2%)
Roseobacter SAGs
PC2 (9.9%)
7am/9pm
r²=0.44
PC1 (45.8%)
Pelagibacter
9am/6pm
r²=0.37
PC2 (16.2%)
PC1 (28.1%)
SAR116 cluster
9am/6pm
r²=0.33
PC2 (7.2%)
PC1 (16.4%)
SAR86 cluster
9am/6pm
r²=0.32
PC2 (8.2%)
PC1 (13.3%)
SAR324 cluster
PC2 (8.2%)
9/9 4pm
r²=0.43
PC1 (12.1%)
9/7 10:38pm
9/8 1:37am
9/8 3:36am
9/8 5:36am
9/8 7:35am
9/8 9:37am
9/8 2:41pm
9/8 5:36pm
9/8 7:37pm
9/8 9:36pm
9/8 11:37pm
9/9 1:36am
9/9 6:53am
9/9 9:35am
9/9 11:35am
9/9 1:35pm
9/9 3:34pm
9/9 5:36pm
9/9 10:36pm
9/10 1:37am
9/10 3:36am
9/10 5:36am
9/10 7:37am
9/10 9:36am
9/10 2:46pm
9/10 5:36pm
9/10 7:36pm
9/10 9:35pm
9/10 11:36pm
9/11 1:36am

335


A B C D E F G
Pro. GA16 3 3 3 0 0 0 0
Pro. GA26 1 1 1 0 0 0 0
Pro. GA4 11 11 0 0 0 0 0
Pro. GA6 2 2 0 0 0 0 0
Pro. GA11 7 7 1 0 0 0 0
Pro. GA15 31 31 4 0 0 1 0
Pro. GA17 64 64 3 0 0 3 0
Pro. GA24 3 3 0 0 0 0 0
Pro. GA41 2 2 0 0 0 0 0
SAR11 GA20 31 14 0 6 0 2 1
Pro. GA45 1 1 0 0 0 0 0
SAR11 GA19 1 1 0 0 0 0 0
Pro. GA27 2 2 1 0 0 1 0
Pro. GA20 3 3 0 0 0 0 0
Pro. GA43 2 2 2 0 0 0 0
Roseo. GA12 60 55 2 7 3 0 3
SAR116 GA5 42 18 0 6 0 2 1
SAR11 GA16 2 2 0 0 0 0 0
Roseo. GA3 17 17 0 0 3 1 0
SAR324 GA3 8 2 0 0 0 0 0
SAR86 GA7 4 0 0 2 0 0 0
SAR11 GA23 1 0 0 0 0 0 0
SAR324 GA9 22 0 0 1 2 3 2
Pro. GA5 14 14 0 0 4 1 0
Pro. GA7 3 3 0 0 0 0 0
Pro. GA9 25 25 0 0 1 2 0
Pro. GA23 9 7 1 0 0 1 1
SAR11 GA18 32 26 0 0 9 5 0
SAR11 GA6 13 12 0 0 5 3 1
SAR116 GA2 14 10 0 0 8 1 1
Roseo. GA8 17 17 0 0 5 5 0
Pro. GA12 65 65 0 0 2 4 1
Pro. GA35 19 19 2 0 2 0 0
SAR11 GA17 8 8 0 1 1 1 0
SAR324 GA6 3 2 0 0 0 0 1
SAR11 GA4 7 4 0 0 0 0 0
SAR86 GA2 9 1 0 0 1 0 0
SAR86 GA5 22 3 0 0 6 0 0
Pro. GA42 2 0 0 0 0 0 0
Pro. GA37 1 0 0 0 0 0 0
SAR11 GA15 1 0 0 0 0 0 0
Pro. GA40 1 0 0 0 0 0 0
SAR11 GA21 1 0 0 0 0 1 0
SAR11 GA13 1 0 0 0 0 0 0
SAR11 GA1 8 1 0 0 0 0 0
SAR86 GA1 2 0 0 0 0 1 0
Roseo. GA1 8 5 0 0 0 0 0
Roseo. GA7 20 7 0 0 0 2 0
SAR116 GA1 16 4 0 0 0 1 0
SAR324 GA1 15 0 0 0 0 0 1
SAR116 GA4 10 2 0 0 0 3 0
Pro. GA1 623 153 12 1 3 30 22
SAR11 GA7 3 0 0 0 0 1 0
SAR11 GA8 6 0 0 1 0 3 1
SAR11 GA9 3 0 0 0 0 0 1
SAR11 GA11 477 13 1 10 3 39 16
SAR11 GA10 4 1 0 0 0 1 1
SAR86 GA6 397 0 5 28 6 34 2
Pro. GA10 2 2 0 0 0 1 0
Pro. GA8 8 8 0 0 0 1 1
Pro. GA13 74 74 0 0 1 11 1
Pro. GA32 3 3 0 0 0 0 0
Pro. GA39 58 58 0 0 0 4 2
Pro. GA34 1 1 0 0 0 0 0
Pro. GA25 1 1 0 0 0 0 0
Pro. GA28 9 9 0 0 0 0 0
Pro. GA29 34 34 0 0 0 4 0
Roseo. GA9 23 21 0 0 0 4 7
SAR116 GA6 2 2 0 0 0 0 0
SAR116 GA7 503 15 4 12 23 71 41
SAR324 GA4 626 1 0 21 15 59 70
Roseo. GA5 8 6 0 0 0 2 1
Roseo. GA6 163 27 0 2 3 13 27
SAR86 GA4 9 0 0 0 2 3 0
SAR324 GA5 3 0 0 0 0 1 1
Pro. GA19 19 18 5 0 0 1 0
Pro. GA30 1 1 1 0 0 0 0
Pro. GA3 2 2 0 0 0 0 0
Pro. GA18 3 3 0 0 0 0 0
Pro. GA31 16 16 7 1 6 0 0
Pro. GA22 93 90 3 2 2 5 0
Pro. GA33 15 15 2 9 2 0 0
Pro. GA36 30 30 0 16 0 0 0
Pro. GA44 2 2 0 1 0 0 0
Pro. GA48 52 52 2 18 1 4 0
SAR11 GA12 1 1 0 0 0 0 0
SAR116 GA9 1 1 0 0 0 0 0
Roseo. GA13 2 1 0 1 0 0 0
SAR324 GA8 3 0 0 0 0 0 0
Pro. GA2 2 2 0 0 0 0 0
Pro. GA46 3 3 0 0 0 0 0
Pro. GA47 18 18 0 0 0 2 1
Roseo. GA11 7 4 0 0 0 0 5
SAR11 GA3 11 5 0 0 0 3 1
SAR116 GA8 17 4 0 0 0 0 9
Pro. GA14 3 3 0 0 0 0 0
Pro. GA21 6 6 0 0 0 1 1
Roseo. GA2 9 9 2 0 0 0 0
Roseo. GA4 18 15 0 0 0 1 1
SAR11 GA22 40 18 0 0 0 4 1
Pro. GA38 1 1 0 0 0 0 0
SAR11 GA14 12 11 0 0 0 1 5
SAR324 GA7 29 1 0 0 1 1 13
SAR11 GA5 19 13 0 0 0 1 5
SAR86 GA8 1 1 0 0 0 0 0
SAR86 GA9 7 4 0 0 0 0 0
SAR86 GA3 31 0 0 0 0 4 0
SAR11 GA2 3 3 0 0 0 0 1
SAR116 GA3 32 4 0 0 0 3 1
SAR324 GA2 5 0 0 0 0 0 4
Roseo. GA10 1 0 0 0 0 0 1
SAR86 GA10 6 0 0 0 0 1 0

336