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].

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

21   **Abstract**: Oscillating diurnal rhythms of gene transcription, metabolic activity and behavior are

22   found in all three domains of life. Diel cycles in naturally occurring heterotrophic bacteria and

23   archaea however, have rarely been observed. Here we report time-resolved whole genome

24   transcriptome profiles of multiple, naturally occurring oceanic bacterial populations sampled *in*

25   *situ* over three days. As anticipated, the cyanobacterial transcriptome exhibited pronounced diel

26   periodicity. Unexpectedly however, several different heterotrophic bacterioplankton groups also

27   displayed diel cycling in many of their gene transcripts. Furthermore, diel oscillations in

28   different heterotrophic bacterial groups suggested population-specific timing of peak transcript

29   expression in a variety of metabolic gene suites. These staggered multispecies waves of diel gene

30   transcription may influence both the tempo and mode of matter and energy transformation in the

31   sea.

32

33

2

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 *Prochlorococus* and several proteorhodopsin-containing or photoheterotrophic bacteria,

56   including members of the *Pelagibacter* (SAR11), *Roseobacter,* SAR116, SAR86, and SAR324

57   clades (Fig. S2).

58   Phylogenetic analysis of gene transcripts in the most abundant taxa revealed the presence

59   of some microdiversity (Figs. S3-8). The most abundant transcripts sampled at any given time

60   point however, were dominated by only a few genotypes within each population that persisted

61   throughout the sampling period.   An exception was *Roseobacter*, with transcripts for two

62   different genes (*groEL* and *dnaK*) indicating the presence of a genotype that started at a very low

63   abundance and increased in representation over the course of the time series. This variability

64   could be due to an injection of a new population as water masses mixed during the latter portion

65   of the time series, or possibly to an alteration in the relative transcriptional activities of two

66   ecotypes that are responding to changes in the surrounding environment.

67   Transcriptional activity in *Prochlorococcus* was highly dependent on the time of day.

68   Harmonic regression analyses indicated that nearly half (1,491) of all *Procholorococcus*

69   population transcripts were significantly periodic (Table 1; Table S3; Fig 1). The expression

70   patterns observed were similar to those of monocultures growing in controlled laboratory settings

71   (*4*) but there were also notable differences (Fig. 1). For example, photosystem I gene expression

72   exhibited a double peak in the wild *Prochlorococcus* transcriptome around noon (Fig. 1). In

73   contrast, under laboratory conditions most photosystem I genes, i.e. *psaL* and *psaF*, were found

74   to peak just before noon, while *psaA* and *psaB* peaked shortly after noon (*4*).

75   The largest discrepancy between *Prochlorococcus* laboratory studies and our field

76   observations was that a considerable number of *Prochlorococcus* transcripts in our field

77   populations peaked around midday (Fig. 1). Some of these genes did exhibit periodicity in

78   cultures, but peaked at a different time of day in field populations. A larger fraction of these

79   mid-day peaking transcripts were either not periodically expressed, or were not present in the

80   culture experiments. In addition, 62% of the 10am - 4pm peaking transcripts in our field study

81   lacked KEGG orthology annotations, as opposed to those peaking in the evening or late at night.

82   A number of factors may be responsible for differences in transcript dynamics between in

83   laboratory cultures versus field *Prochlorococcus* populations. Maximal light levels at our study

84   site at 23 m depth were frequently two fold higher (450 umol $Q/m^{-2}/s^{-1}$) than those used in

85   laboratory microarray experiments (232 umol $Q/m^{-2}/s^{-1}$) (*4*). Fundamental genetic differences

86   between our field populations and the *Prochlorococcus* strain used in laboratory culture

87   experiments likely also contribute to the differences we observed. Other variables, including

88   nutrient composition and organismal interactions, may also be a factor in the observed

89   differences. While we could not identify obvious trends in the type or function of transcripts

90   showing peak expression during the mid-day period, they did include a wide range of enzymatic

91   functions that are more consistent with nutrient-responsive metabolic changes rather than a

92   simple high-light stress response.

93   An abundant *Roseobacter* population also showed strong diel oscillations in its

94   transcriptome profile, most notably in expressed genes involved in bacteriochlorophyll-

95   associated aerobic anoxygenic photosynthesis (AAnP). Overall, a large fraction of *Roseobacter*

96   transcripts were periodically expressed (Table 1). Of these, the majority peaked during daylight

97   hours, with only a few gene transcripts peaking at night (Fig. 2). While this pattern contrasts

98   with that observed in *Prochlorococcus*, where most diurnally regulated transcripts peaked at

99   dawn or dusk, it was consistent with transcriptional regulation recently reported in

100 *Dinoroseobacter shibae* (*16*).

101   Thirty five of the forty significantly periodic *Roseobacter* transcripts that peaked between

102   11 pm and 7 am encoded genes belonging to a large photosynthetic "superoperon" (Fig. S9).

103   Nightly expression of these genes, followed by immediate repression upon light onset, is

104   consistent with the *D. shibae* study (*16*), and may be preparing cells for efficient solar energy

105   harvest in the early morning hours. Functions that peaked during the day-time hours included

106   ribosomal proteins, respiratory transcripts, genes involved in amino acid metabolism, and

107   transporters (Fig. 2).

108   Proteorhodopsin-containing photoheterotrophs including members of the SAR11,

109   SAR116, and SAR86 also showed evidence of diel periodicity in many of their gene transcripts

110   (Fig. 2, Table 1). Interestingly, all opsin-containing bacteria analyzed (SAR11, SAR116, SAR86,

111   and SAR324) exhibited statistically significant diel oscillations in their proteorhodopsin gene

112   transcripts (Table S3; Fig. S10). Peak expression of the opsin transcripts occurred near dawn in

113   all these populations (Fig. S10), potentially optimizing solar energy capture by the light-driven,

114   proton-pumping rhodopsins.

115   Principal components analysis distinguished time series samples for each heterotroph by

116   time of day (Fig. 3) and showed significant correlation with the light-driven behavior of

117   *Prochlorococcus* (Table 1). Overall, this data is consistent with profound, genome-wide

118   transcriptional changes across the day-night cycle for each population. In addition, co-clustering

119   of transcripts using GeneARMA (GA; (*14, 17*)) revealed suites of gene transcripts that exhibited

120   similar expression patterns among different taxa (Fig. 4; Fig. S11-14; Table S4). For example, a

121   group of transcripts that that fit highly similar GA expression models across multiple species

122   included Pro GA5, Pro GA7, Pro GA9, Pro GA23, SAR11 GA6. SAR11 GA18, SAR116 GA2,

123   and *Roseobacter* GA8 (Fig 4; Fig S14; Table S4). These multispecies, day-peaking transcripts

124   (Fig S14; Table S4) included gene products associated with respiration (*Procholorococcus*,

125   SAR11, SAR116, *Roseobacter*), nitrogen metabolism (*Procholorococcus*, SAR11, SAR116),

126   glycine metabolism (*Procholorococcus*, SAR11, *Roseobacter*), carbon monoxide metabolism

127   (SAR116, *Roseobacter*) and DNA synthesis (*Procholorococcus*, *Roseobacter*).   This co-

128   clustering of gene transcripts reveals a complex pattern of expression through the day and across

129   the time series, and provides evidence for parallel trends in gene expression across multiple

130   species (Fig. 4, Fig. S11-14, Table S4).

131   Together, the transcriptional profiles of *Roseobacter*, SAR11, SAR116 and SAR86

132   indicate diel cycling of metabolic gene transcripts, and suggest a multispecies wave-like

133   progression of upregulated gene suites across the day/night cycle (Fig. 4). Most conspicuously, a

134   regular diel succession of translational, transcriptional and respiratory gene transcripts was

135   followed by peaks in transporter transcripts that possibly reflect a metabolic recovery phase (Fig.

136   2). Many of these metabolic pathway transcripts peaked earlier in the day in *Roseobacter* field

137   populations relative to other bacterial heterotrophs (Fig. 2, Fig 4, Table S4).

138   The overall transcriptional profile of SAR324 did not show as many transcript diel

139   oscillations as other heterotrophic taxa.   Instead, principal components analysis clustered

140   SAR324 transcripts according to the day that they were collected (Fig. 3). In particular, the

141   SAR324 group showed a strong separation between the first portion of the time series and the

142   second in principal components analysis (Fig. 3).   This split appears to be associated with the

143   increases in temperature and salinity observed across the time series (Fig. S1).

144   The diurnal patterns reported here for open ocean heterotrophic bacterioplankton were

145   different from those observed in a previous study of phylogenetically related coastal

146   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

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

171

172   **References and Notes:**

173   1.   D. Bell-Pedersen *et al.*, Circadian rhythms from multiple oscillators: lessons from diverse organisms.
174   *Nature reviews. Genetics* **6**, 544 (2005).
175   2.   I. M. Axmann, S. Hertel, A. Wiegard, A. K. Dorrich, A. Wilde, Diversity of KaiC-based timing systems in
176   marine Cyanobacteria. *Marine genomics* **In press,** http://dx.doi.org/10.1016/j.margen.2013.12.006,
177   (2014).
178   3.   H. Ito *et al.*, Cyanobacterial daily life with Kai-based circadian and diurnal genome-wide transcriptional
179   control in *Synechococcus elongatus*. *Proc. Natl. Acad. Sci. U.S.A.* **106**, 14168 (2009).
180   4.   E. R. Zinser *et al.*, Choreography of the transcriptome, photophysiology, and cell cycle of a minimal
181   photoautotroph, prochlorococcus. *PloS one* **4**, e5135 (2009).
182   5.   K. Whitehead, M. Pan, K. Masumura, R. Bonneau, N. S. Baliga, Diurnally entrained anticipatory behavior
183   in archaea. *PloS one* **4**, e5485 (2009).
184   6.   J. F. Flint, D. Drzymalski, W. L. Montgomery, G. Southam, E. R. Angert, Nocturnal production of
185   endospores in natural populations of *Epulopiscium*-like surgeonfish symbionts. *Journal of bacteriology*
186   **187**, 7460 (2005).
187   7.   A. M. Wier *et al.*, Transcriptional patterns in both host and bacterium underlie a daily rhythm of anatomical
188   and metabolic change in a beneficial symbiosis. *Proc. Natl. Acad. Sci. U.S.A.* **107**, 2259 (2010).
189   8.   J. M. Gasol *et al.*, Diel variations in bacterial heterotrophic activity and growth in the northwestern
190   Mediterranean Sea. *Mar Ecol Prog Ser* **164**, 107 (1998).
191   9.   C. Winter, G. J. Herndl, M. G. Weinbauer, Diel cycles in viral infection of bacterioplankton in the North
192   Sea. *Aquat Microb Ecol* **35**, 207 (2004).
193   10.   M. Galí *et al.*, Diel patterns of oceanic dimethylsulfide (DMS) cycling: Microbial and physical drivers.
194   *Global Biogeochem Cycles* **27**, 620 (2013).
195   11.   S. M. Gifford, S. Sharma, M. Booth, M. A. Moran, Expression patterns reveal niche diversification in a
196   marine microbial assemblage. *The ISME journal* **7**, 281 (2013).
197   12.   E. A. Ottesen *et al.*, Pattern and synchrony of gene expression among sympatric marine microbial
198   populations. *Proc. Natl. Acad. Sci. U.S.A.* **110**, E488 (2013).
199   13.   E. A. Ottesen *et al.*, Metatranscriptomic analysis of autonomously collected and preserved marine
200   bacterioplankton. *The ISME journal* **5**, 1881 (2011).
201   14.   Materials and methods are available as supplementary material on Science Online
202   15.   C. M. Preston *et al.*, Underwater application of quantitative PCR on an ocean mooring. *PloS one* **6**, e22522
203   (2011).
204   16.   J. Tomasch, R. Gohl, B. Bunk, M. S. Diez, I. Wagner-Döbler, Transcriptional response of the
205   photoheterotrophic marine bacterium *Dinoroseobacter shibae* to changing light regimes. *The ISME journal*
206   **5**, 1957 (2011).
207   17.   N. Li *et al.*, Functional clustering of periodic transcriptional profiles through ARMA(p,q). *PloS one* **5**,
208   e9894 (2010).
209   18.   L. Steindler, M. S. Schwalbach, D. P. Smith, F. Chan, S. J. Giovannoni, Energy starved Candidatus
210   Pelagibacter ubique substitutes light-mediated ATP production for endogenous carbon respiration. *PloS*
211   *one* **6**, e19725 (2011).
212   19.   Z. Liu, W. Hsiao, B. L. Cantarel, E. F. Drábek, C. Fraser-Liggett, Sparse distance-based learning for
213   simultaneous multiclass classification and feature selection of metagenomic data. *Bioinformatics* **27**, 3242
214   (2011).
215   20.   J. Okansen et al. Package "vegan": Community ecology package.
216   http://cran.rproject.org/web/packages/vegan//vegan.pdf (2012).
217   21.   J. C. Robidart *et al.*, Ecogenomic sensor reveals controls on N-fixing microorganisms in the North Pacific
218   Ocean. *The ISME journal* doi: 10.1038/ismej.2013.244, (2014).
219   22.   F. J. Stewart, E. A. Ottesen, E. F. DeLong, Development and quantitative analyses of a universal rRNA-
220   subtraction protocol for microbial metatranscriptomics. *The ISME journal* **4**, 896 (2010).

221 23. Y. Shi, G. W. Tyson, J. M. Eppley, E. F. DeLong, Integrated metatranscriptomic and metagenomic
222 analyses of stratified microbial assemblages in the open ocean. *The ISME journal* **5**, 999 (2011).
223 24. S. Rodrigue *et al.*, Unlocking short read sequencing for metagenomics. *PloS one* **5**, e11840 (2010).
224 25. S. M. Kielbasa, R. Wan, K. Sato, P. Horton, M. C. Frith, Adaptive seeds tame genomic sequence
225 comparison. *Genome research* **21**, 487 (2011).
226 26. B. K. Swan *et al.*, Prevalent genome streamlining and latitudinal divergence of planktonic bacteria in the
227 surface ocean. *Proc. Natl. Acad. Sci. U.S.A.* **110**, 11463 (2013).
228 27. R. J. Newton *et al.*, Genome characteristics of a generalist marine bacterial lineage. *The ISME journal* **4**,
229 784 (2010).
230 28. M. Kanehisa, S. Goto, KEGG: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* **28**, 27
231 (2000).
232 29. Team RDC (2012) R: A Language and Environment for Statistical Computing (R Foundation for Statistical
233 Computing, Vienna).
234 30. Y. Benjamini, Y. Hochberg, Controlling the false discovery rate - a practical and powerful approach to
235 multiple testing. *J Roy Stat Soc B Met* **57**, 289 (1995).
236 31. J. P. Baudry, A. E. Raftery, G. Celeux, K. Lo, R. Gottardo, Combining Mixture Components for Clustering.
237 *Journal of computational and graphical statistics : a joint publication of American Statistical Association,*
238 *Institute of Mathematical Statistics, Interface Foundation of North America* **9**, 332 (2010).
239 32. R. C. Edgar, Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* **26**, 2460
240 (2010).
241 33. R. C. Edgar, MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic*
242 *Acids Res.* **32**, 1792 (2004).
243 34. M. N. Price, P. S. Dehal, A. P. Arkin, FastTree: computing large minimum evolution trees with profiles
244 instead of a distance matrix. *Molecular biology and evolution* **26**, 1641 (2009).
245 35. J. G. Caporaso *et al.*, QIIME allows analysis of high-throughput community sequencing data. *Nature*
246 *methods* **7**, 335 (2010).

247

248

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

260

261

268

269

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

285

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

298

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

301 and arcsin transformed to approximate normality (*19*). Symbol color denotes time of day and
302 shape denotes day of collection. Grey lines connect samples to centroids for selected sample
303 groupings that separate points well. *Roseobacter* SAGs: samples collected between 7am and 9pm
304 (vs. 9pm to 7am); SAR11, SAR116 and SAR86 cluster, samples collected between 9am and 6pm
305 (and vice versa); SAR324 cluster, samples collected before or after 9/9 4pm. All factor
306 correlations shown were highly significant (p = 0.001). Alternative time of day categories were
307 also highly significant for Roseobacter SAGs, SAR11, SAR116 and SAR86. SAR116 (r^2 0.10, p
308 0.037) and SAR86 (r^2 0.12, p 0.019) also correlated weakly with the grouping shown for
309 SAR324. All analyses carried out using functions in the vegan software package (*20*).

310

311 **Fig. 4**. Timing of expression of functional gene clusters in different taxa clustered by similarity.
312 Heatmap shows cluster models for all geneARMA clusters, colored by mean-centered relative
313 expression (red=high, blue=low). Black and yellow bars show the daily photoperiod. Each box
314 represents a single sampling event, for sample times see Table S1. Dendrograms show cluster
315 model similarity (Pearson correlations, average linkage clustering, scale bar at upper right
316 represents a correlation of 0.5). The total number of genes (A), significantly periodic genes (B),
317 and genes associated with Photosynthesis (C), Ribosome (D), Oxidative Phosphoryloation (E),
318 Amino Acid Metabolism (F), and Transport (G) (defined as for Figure 2), are listed for each
319 cluster.
320

321 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)

322 [1] The total number of sequence reads assigned to each taxon bin
323 [2] The total number of unique ortholog clusters (see Database S1) with at least one mapped sequence.
324 [3] The total number of sequences identified as showing 24-hour periodicity using harmonic regression.
325 [4] Proportion of variance explained by 24-hour periodicity in constrained principal components analysis.
326 [5] Procrustes correlation between the first two principal components from *Prochlorococcus* and other taxa
327   (unconstrained principal components analysis as shown in Fig. 3). P-value based on 999 permutations.
328 [6] Correlation between pairwise sample similarities from *Prochlorococcus* and heterotrophic taxa based on Mantel
329   test on Euclidean distance matrices. P-value based on 999 permutations.
330
331

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

Label	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	1	0	0	0
SAR86 GA5	22	3	0	0	0	6	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	0
SAR116 GA4	10	2	0	0	0	3	0
Pro. GA1	623	153	12	1	3	30	22
SAR11 GA7	3	0	0	1	0	3	1
SAR11 GA8	6	0	0	0	0	0	0
SAR11 GA9	3	0	0	0	0	0	0
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	1
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	1	1
Pro. GA21	6	6	0	0	0	1	1
Roseo. GA2	9	9	2	0	0	0	0
Roseo. GA4	18	15	0	0	2	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	0
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