Pattern and synchrony of gene expression among sympatric marine microbial populations

Elizabeth A. Ottesen, Curtis R. Young, John M. Eppley, John P. Ryan, Francisco P. Chavez, Christopher A. Scholin, and Edward F. DeLong

Planktonic marine microbes live in dynamic habitats that demand rapid sensing and response to periodic as well as stochastic environmental change. The kinetics, regularity, and specificity of microbial responses in situ, however, are not well-described. We report here simultaneous multilocus genome-wide transcriptome profiling in a naturally occurring picoplankton community. An in situ robotic sampler using a Lagrangian sampling strategy enabled continuous tracking and repeated sampling of coherent microbial populations over 2 d. Subsequent RNA sequencing analyses yielded genome-wide transcriptome profiles of eukaryotic (Ostreococcus) and bacterial (Synechococcus) photosynthetic picoplankton as well as proteorhodopsin-containing heterotrophs, including Pelagibacter, SAR86-cluster Gamma proteobacteria, and marine Euryarchaea. The photosynthetic picoplankton exhibited strong diel rhythms over thousands of gene transcripts that were remarkably consistent with diel cycling observed in laboratory pure cultures. In contrast, the heterotrophs did not cycle diurnally. Instead, heterotrophic picoplankton populations exhibited cross-species synchronous, tightly regulated, temporally variable patterns of gene expression for many genes, particularly those genes associated with growth and nutrient acquisition. This multilocus, population-wide gene regulation seemed to reflect sporadic, short-term, reversible responses to high-frequency environmental variability. Although the timing of the environmental responses among different heterotrophic species seemed synchronous, the specific metabolic genes that were expressed varied from taxon to taxon. In aggregate, these results provide insights into the kinetics, diversity, and functional patterns of microbial community response to environmental change. Our results also suggest a means by which complex multispecies metabolic processes could be coordinated, facilitating the regulation of matter and energy processing in a dynamically changing environment.

Results and Discussion

Marine surface water microbes were collected and preserved by a robotic system (10) suspended beneath a free-drifting buoy deployed off the coast of northern California (Fig. 1). Over a 2-d sampling period, the instrument drifted 50.3 km along the warm side of a front that was generated by coastal upwelling to the east (Fig. 1). Samples for community RNA sequencing were collected and preserved every 4 h. Portions of the sampling track were marked by strong vertical gradients in salinity and temperature (Fig. 1 C and D). However, current velocity data suggest that the water sampled for microbial transcriptomic analysis was relatively stable horizontally (SI Appendix, Fig. S1 and Table S1). The taxonomic representation in the metatranscriptome sampled over

Significance

Microbial communities regulate the cycling of energy and matter in the environment, yet how they respond to environmental change is not well-known. We describe here a day in the life of wild planktonic microbial species using robotic sampling coupled with genome-wide gene expression analysis. Our results showed that closely related populations, as well as very different bacterial and archaeal species, displayed remarkably similar time-variable synchronous patterns of gene expression over 2 d. Our results suggest that specific environmental cues may elicit cross-species coordination of gene expression among diverse microbial groups, potentially enabling multispecies coupling of metabolic activity.

Author contributions: E.A.O., C.A.S., and E.F.D. designed research; E.A.O. performed research; J.M.E. and C.A.S. contributed new reagents/analytic tools; E.A.O., C.R.Y., J.M.E., J.P.R., F.P.C., C.A.S., and E.F.D. analyzed data; and E.A.O. and E.F.D. wrote the paper.

www.pnas.org/cgi/doi/10.1073/pnas.1222099110

PNAS PLUS

MICROBIOLOGY

ENVIRONMENTAL SCIENCES

this time also remained relatively constant (Fig. 1). Of the over 2.4 million sequences that were assigned to 8,117 unique National Center for Biotechnology Information (NCBI) Taxonomy IDs, ~98% matched taxa that were detected in all 13 samples. Additionally, our samples showed greater overall similarity in taxonomic composition to one another across all of the time points than did samples collected over a 24-h period at a fixed spatial location in Monterey Bay (6) (SI Appendix, Fig. S2).

Our analyses focused on transcriptional dynamics among five abundant microbial populations in the sampled community, including Ostreococcus, Synechococcus, Pelagibacter, SAR86 cluster Gammaproteobacteria (SAR86), and marine group II Euryarchaea (MGII) (Table 1). These populations represent widely distributed and ecologically important clades of marine picoplankton (11–15).

Transcripts mapping to each of these groups were identified and annotated using a newly developed computational workflow (SI Appendix, Fig. S3) that assigned sequences to specific taxon bins based on best-scoring matches within the full NCBI database. Within each taxon bin, transcript counts for genes shared between multiple reference genomes of the same taxon were combined, and analyses of transcriptional dynamics focused on changes in relative transcript abundance within each specific taxonomic population, independent of fluctuations in its abundance relative to the total community transcriptome. (Table 1 and SI Appendix, Figs. S3–S11 and Tables S2 and S3 have details of transcript mapping and annotation.)

To confirm that complex transcriptional dynamics could be observed within transcriptional profiles extracted from community-wide gene expression datasets, we used cluster analysis to assess global transcriptional patterns among the five most highly represented taxonomic groups (Fig. 2 and SI Appendix, Figs. S12–S17). Within each taxon, we identified groups of genes that shared similar transcriptional profiles using the geneARMA software package (16), which uses an autoregressive moving average model, ARMA (p,q), for the longitudinal covariance structure and Fourier series functions to model gene expression patterns. As anticipated, a large number of coexpressed genes with apparent 24-h periodicity were identified among Ostreococcus and Synechococcus populations. Additionally, principal components analysis clearly separated the transcriptome profiles of these taxa based on time of day (Fig. 2B). Pure cultures of both Ostreococcus and Synechococcus are known to have fully functional circadian clocks that coordinate large-scale transcriptional dynamics (17, 18), and those rhythms were readily

![Fig. 1. Sampling locations and sample characteristics. (A) The ESP drift track imposed over a map showing average sea surface temperature (Polar orbiting environment satellites, advanced very high resolution radiometer, local area coverage. Western United States, day/night, 5 x 5-pixel (5 x 5-km) median-filtered composite from September 21, 2010; cloud cover precluded satellite observation during the sampling period). Inset shows the sampling location relative to western North America. (B) Transcript abundances of major taxa represented as percent of sequences with matches in the NCBI nr peptide database. (C) Environmental conditions in the immediate vicinity of the ESP (measurements taken by ESP-mounted instruments). Grey bars represent sample collection times. (D) Integrated depth profile showing salinity gradients surrounding the sampling location (dotted line). Measurements were taken near the ESP by a ship-deployed instrument. The arrow in A indicates the direction of the drift.]
some, and Calvin cycle transcripts largely exhibited a morning peak in expression, whereas cytochrome oxidase transcripts peaked in the evening. In contrast to observed patterns among *Ostreococcus*, only one photosystem gene (Cluster 8842 psaK) and relatively few antenna proteins (4 of 25) were identified as significantly periodic in our dataset. Genes in these categories did, however, show high-amplitude and highly periodic expression, suggesting that they may be differentially regulated based on environmental factors. Also unlike *Ostreococcus*, evidence for diel periodicity in growth and division among this *Synechococcus* population was relatively weak, with no cell cycle or DNA replication transcripts showing significantly periodic expression. However, given the complicated relationship between nutrient status and the timing of DNA replication and cell division in *Synechococcus* (19, 20), the unknown growth state of this population, and the apparent presence in our dataset of two ecologically distinct *Synechococcus* clades (*SI Appendix, Fig. S8*), a lack of a clear periodic signal in population-averaged transcript abundance for individual cell cycle and DNA replication genes is perhaps not surprising.

Both the *Ostreococcus* and *Synechococcus* data reflected broad trends previously observed in laboratory monocultures. The diurnal timing of expression of transcripts in wild *Ostreococcus* populations in particular was remarkably consistent with gene expression patterns observed in microarray-based laboratory studies of *O. tauri* cultures grown in 12:12-h light:dark cycles (18) (Fig. 4). However, there were also a number of differences between results obtained for our natural populations and those results observed in laboratory analyses of pure cultures.

A direct comparison of our field study with any previously published laboratory studies of *Synechococcus* was problematic, because most existing datasets (17) focus on *S. elongatus*, a fresh-water species, and were performed under constant light illumination. It is worth noting, however, that our study identified fewer *Synechococcus* transcripts that exhibited periodic trends in expression compared with laboratory studies of the freshwater *Synechococcus* species. However, the orthologs identified in our field populations do not simply represent a high-amplitude subset of the periodically expressed transcripts detected in laboratory studies. Of 69 periodically expressed *Synechococcus* orthologs in our field study that could be mapped to probes in the microarray, evidence for diel periodicity in growth and division for this *Ostreococcus* population was relatively weak, with no cell cycle or DNA replication transcripts showing significantly periodic expression. However, given the complicated relationship between nutrient status and the timing of DNA replication and cell division in *Synechococcus* (19, 20), the unknown growth state of this population, and the apparent presence in our dataset of two ecologically distinct *Synechococcus* clades (*SI Appendix, Fig. S8*), a lack of a clear periodic signal in population-averaged transcript abundance for individual cell cycle and DNA replication genes is perhaps not surprising.

Both the *Ostreococcus* and *Synechococcus* data reflected broad trends previously observed in laboratory monocultures. The diurnal timing of expression of transcripts in wild *Ostreococcus* populations in particular was remarkably consistent with gene expression patterns observed in microarray-based laboratory studies of *O. tauri* cultures grown in 12:12-h light:dark cycles (18) (Fig. 4). However, there were also a number of differences between results obtained for our natural populations and those results observed in laboratory analyses of pure cultures.

A direct comparison of our field study with any previously published laboratory studies of *Synechococcus* was problematic, because most existing datasets (17) focus on *S. elongatus*, a fresh-water species, and were performed under constant light illumination. It is worth noting, however, that our study identified fewer *Synechococcus* transcripts that exhibited periodic trends in expression compared with laboratory studies of the freshwater *Synechococcus* species. However, the orthologs identified in our field populations do not simply represent a high-amplitude subset of the periodically expressed transcripts detected in laboratory studies. Of 69 periodically expressed *Synechococcus* orthologs in our field study that could be mapped to probes in the laboratory-based microarray study, 24 orthologs were not identified as significantly periodic under laboratory conditions.

For *Ostreococcus*, the populations reported here and a previous laboratory study (18) identified ~2,000 periodically expressed
Ostreococcus transcripts. However, of 1,683 significantly periodic orthologs in our field populations that could be mapped to probes in the O. tauri microarray, 881 orthologs were not identified as significantly periodic in the laboratory study. Reprocessing the laboratory microarray data using our regression-based approach (with a Gaussian error model) increased the overlap in significantly periodic genes, but this analysis still yielded 393 genes identified as periodic in our field data but not in the laboratory. We did not identify any obvious biological trends among Ostreococcus transcripts that were identified as periodically expressed in the field but not the laboratory. Although some of these differences in gene expression patterns may be the result of methodological differences, many are likely to represent responses to cues present in the natural environment but not within the relatively static laboratory environment.

In sum, these analyses validate our approach and confirm that complex transcriptional patterns within distinct populations can be resolved within bulk community RNA profiles. Although previous studies have suggested single-time point day/night differences in the overall transcriptional profiles of marine microbial communities, our analyses here provide a much higher-resolution picture of genome-wide diel transcriptional dynamics among different microbial populations in a natural microbial community in situ.

**Transcriptional Dynamics Within Pelagibacter Population Transcripts.** The naturally occurring Pelagibacter populations that we sampled did not exhibit strong circadian rhythms of gene expression. We did however observe evidence for well-orchestrated, genome-wide transcriptional regulation within this group. Hierarchical clustering of samples and pathways showed a large degree of covariance between some major metabolic pathways (Fig. 5A). In particular, the pathway-level signal for ribosomal proteins and oxidative phosphorylation showed strong positive correlation with one another (correlation coefficient = 0.98, P value = 1 × 10⁻⁸) and were negatively correlated with many transport gene transcripts, including the ATP binding cassette (ABC) transporter.

**Fig. 2.** Global transcriptional profiles from phototrophic and heterotrophic taxon bins. Heat maps depict third-order Fourier series models from geneARMA clustered transcripts within phototrophic (A) and heterotrophic (C) taxa. Heat maps show model amplitude at the 13 time points (columns) for each of the geneARMA clusters (rows). GeneARMA cluster models were normalized to an amplitude of one. Membership information and individual gene transcript traces are shown in SI Appendix, Figs. S13–S17. (B) Principal components analysis of Ostreococcus and Synechococcus transcriptomes. Axes represent the first and second principal components and are labeled with the percent of total variance explained.
family (correlation coefficient = −0.88, P value = 1 × 10^{-5} for ribosomal proteins vs. ABC). Principal components analysis suggested that these metabolic signals explain more of the variability observed in wild *Pelagibacter* transcript profiles than any of the measured environmental parameters (Fig. 5B). Furthermore, a Poisson regression-based analysis found that 101 of 1,810 observed *Pelagibacter* transcripts were significantly correlated with pathway-level signals for either ribosome biosynthesis or ABC transport (*SI Appendix, Tables S4 and S5*). A total of 74 of these transcripts was identified as up-regulated in tandem with the ribosome synthesis pathway (and down-regulated with transport up-regulation), including not only transcripts coding for ribosomal proteins but also genes associated with C1 metabolism (21), secretion, ATP synthase (six of nine subunits), and proton-translocating pyrophosphatase. The 27 transcripts that followed the opposite trend (up-regulated with transporters and down-regulated with ribosome synthesis) seemed to represent a generalized transport signal, encompassing not only ABC transporters of amino acids, polyamines, and phosphonates but also TRAP (Tripartite ATP-independent Periplasmic) transporters of carboxylic acids, an ammonium transporter, and an Na^{+}/solute symporter.

In many microbial species, the abundance of ribosomal proteins and their transcripts is tightly regulated with respect to cellular growth rate (22–24), and the relative abundance of ribosomal protein transcripts for a given taxon has been proposed as a metric for assessing in situ growth rates (8). The trends in *Pelagibacter* gene expression that we report here, however, suggest a very dynamic and rapidly fluctuating reallocation of cellular resources between growth and nutrient acquisition. In this context, cell populations exhibiting decreased ribosomal protein synthesis and increased transporter activity are most likely indicators of sporadically limiting substrate availability in the ambient environment. The broad range of transporters that were expressed when ribosomal synthesis was down-regulated does not suggest limitation by any single nutrient. Additionally, neither set of genes seems to reflect stationary-phase responses previously reported in laboratory cultures of *Candidatus P. ubique* (25) (*SI Appendix, Tables S4 and S5*), suggesting that none of these populations have entered a starvation state. Instead, these trends reflect highly dynamic and variable transcriptional responses (and potentially, metabolic and growth rate variability) over short time scales, that seem to be dictated by surrounding environmental and nutrient variability.
Synchronous Transcriptional Dynamics Within *Pelagibacter*, SAR86, and MGII. Although *Pelagibacter* exhibited the strongest and most coherent patterns in transcript abundance among the three proteorhodopsin, synchronous heterotrophic populations examined, SAR86 and MGII populations also exhibited transcriptional changes over the 2–d time series. Interestingly, these patterns seemed to reflect synchronous responses to the same cryptic environmental changes that appeared to be driving *Pelagibacter* transcription dynamics. The degree of similarity between samples for the three organisms was significantly related (Fig. 6A), suggesting that the overall transcriptional profiles of these groups were changing simultaneously, or nearly so. This relationship was even more evident when Procrustes tests were used to compare only the first two axes of population-specific principal components analyses (Fig. 6B), restricting the comparison to the strongest trends in sample-to-sample variability. Furthermore, the relative abundance of transcripts involved in ribosome biosynthesis and oxidative phosphorylation was positively correlated across the 13 time points for the three groups (Fig. 6C and SI Appendix, Table S6). Similarly, independent geneARMA analyses of the three taxa identified gene cluster models exhibiting similar trends in transcript abundance across the three datasets (SI Appendix, Fig. S18). Altogether, these analyses suggest that all three populations were responding to a common environmental signal, exhibiting global synchronous changes in taxon-level taxonomic profiles.

Among heterotrophic marine picoplankton, *Pelagibacter*, SAR86, and MGII have been hypothesized to catabolize different types of carbon molecules (14, 15, 21, 26). *Pelagibacter* seems to use simple peptides, amino acids, osmolytes, and single carbon compounds, whereas the SAR86 and MGII groups have been hypothesized to specialize in the consumption of larger, more complex polymers, such as proteins, polysaccharides, and lipids. Consistent with these predictions, the most abundant transcripts for each of the taxon bins reflected very different metabolic profiles (SI Appendix, Fig. S19). *Pelagibacter* expressed ABC transporters for small peptides, osmolytes, and dicarboxylic acids at high levels, consistent with genomic analyses (26). SAR86 transcripts included a large number of TonB-dependent receptors, previously hypothesized to mediate uptake and metabolism of large polysaccharides and lipids in this organism (14). Finally, the MGII transcriptome was dominated by large cell surface proteins and amino acid transporters, consistent with a hypothesized ability to metabolize large proteins (15). Therefore, although these three very different populations exhibited similar trends in expression of pathways involved in growth and energy metabolism, they did not show similar trends in most metabolic pathways (SI Appendix, Table S6). This trend suggests that the synchronous transcriptional dynamics observed for these groups may reflect bulk changes in the environment.
availability of a broad range of carbon-based substrates rather than responses to the availability of a single common nutrient or substrate limitation.

Overall, the Pelagibacter population showed a stronger global transcriptional response, involving a larger number of transcripts, than the transcriptional responses observed for the SAR86 and MGII populations. Although some of this difference may be due to the higher sequence coverage of Pelagibacter (because of its smaller genome and greater overall abundance), SAR86 and MGII showed less sample-to-sample variation between transcriptional profiles than did Pelagibacter, even when all datasets were resampled to represent an even coverage level. Pelagibacter species have been shown to have highly streamlined genomes with reduced regulatory machinery (26). As a result, these r- proteobacteria may exhibit a smaller range of transcriptional dynamics in response to complex environmental cues, resulting in strong global transcriptional dynamics. In contrast, SAR86 and MGII, with larger genomes and corresponding increased metabolic and regulatory versatility, might be expected to exhibit more complex time- and location-specific behaviors not as easily distinguished using cluster- and correlation-based approaches. Alternatively, the SAR86 and MGII populations may simply respond less strongly to the environmental cues that elicited strong Pelagibacter responses. It is also possible that higher-molecular weight, more-complex substrates preferred by SAR6 and MGII had a more patchy distribution than the simple peptides and osmolytes used by Pelagibacter. As a result, gene expression in SAR86 and MGII populations might be expected exhibit a larger degree of cell-to-cell variability, resulting in weaker or less synchronized transcriptional dynamics when averaged at the temporal (~40 min) and spatial (~1 L) scale of our sample collections.

Notably, although many pathway-level signals from SAR86 and MGII populations were individually correlated with pathway-level signals from Pelagibacter, these populations showed significantly less congruence to each other (SI Appendix, Table S6). The metabolic profiles of SAR86 and MGII, although indicative of different substrate preferences, share the commonality of the binding and hydrolysis of large polymeric substrates, such as cell wall and membrane components. The overall abundances of SAR86 and MGII transcripts in the metatranscriptomes are negatively correlated (correlation coefficient = −0.55, P value = 0.047), which may suggest some degree of niche overlap and competition between these organisms. In contrast, Pelagibacter specializes in a very different fraction of the substrate pool, including low-molecular weight monomers, such as carboxylic acids and amino acids, that are likely to be generated as a byproduct of both SAR86 and MGII metabolic activities.

Altogether, our results revealed unexpected interspecies synchronicity in the regulation of some pathways, as well as a surprising degree of heterogeneity in the transcriptional profiles among photoheterotrophic picoplankton populations. Transcripts encoding ribosomal proteins and genes involved in oxidative phosphorylation were previously identified as highly variable between samples collected at distant geographic locations (5). Notably, we observed as much variability in transcript abundance in these pathways over only a few hours time and in the same water mass, as had been previously reported in transoceanic metabolic surveys. These data suggest that activity levels and respiration rates for heterotrophic populations may be spatially and temporally patchy in marine surface waters, potentially due to episodic substrate releases, such as small-scale lytic events and other stochastic environmental processes. Additional exploration of these behavioral patterns and the environmental cues that control them is likely to provide significant insight into niche specialization of key microbial groups in the planktonic environment.

**Implications.** Episodic environmental variation and subsequent microbial responses play significant roles in shaping marine biogeochemical cycles (1). To better understand and predict microbial community responses to such events, it is critical to observe them on relevant temporal and spatial scales in situ. Here, we
show that Lagrangian sampling combined with microbial community transcriptome analyses can resolve microbial dynamics on time scales of hours to days, yielding robust transcriptional expression patterns. Within a given taxon, the presence of reproducible temporal patterns in the genome-wide transcription profile indicated that a large fraction of individual cells within a given population was responding synchronously, at least within the temporal resolution of our measurements. Furthermore, disparate heterotrophic taxa within the community also seemed to be simultaneously responding to similar environmental cues but expressed different functional gene suites in response to them, suggesting a potential means by which multispecies metabolic and biogeochemical processing might be coordinated.

The specific environmental factors that influence the observed synchronized transcriptional regulation of diverse heterotrophic microbial species are unknown at present. It may be that each species population is responding independently to the same (or simultaneously occurring) physicochemical environmental cues. However, we cannot rule out that these transcriptional patterns are partly influenced by species-to-species communication and signaling cascade events. It is well-known that specific auto-inducer molecules can elicit complex regulatory responses within and between disparate bacterial species (27). It, therefore, seems possible that specific physicochemical environmental cues might be sensed simultaneously by one or a few species. These cues might then be indirectly broadcast to other species by small-molecule signaling, thereby transmitting the response to other community members. Future work, using higher-frequency sampling of microbial community transcriptional profiles, may provide the temporal resolution necessary to distinguish between these different cross-community sensing and response modalities. Regardless of the specific mechanism(s) of multipopulation environmental sensing and response, both the above possibilities could elicit the multispecies transcriptional events that we observed. This temporal entrainment could conceivably serve to coordinate downstream biogeochemical processing and nutrient regeneration. For example, hydrolysis of higher-molecular weight organic compounds by SAR86 and GII *Euryarchaeaa* could produce monomers that were subsequently processed by *Pelagibacter*.

Very little is known about the actual metabolic rates of specific heterotrophic picoplankton species in the environment. Most in situ growth measurements have been derived from bulk-based radioisotope incorporation into DNA or protein across entire assemblages. Complex predator–prey dynamics involving phages and protists further complicate the measurement of species-specific growth rates in situ. The transcriptional profiles of heterotrophic marine picoplankton that we observed suggested that disparate species were responding rapidly to environmental variability with frequent and synchronized up- or down-regulation of transcripts in many pathways. In particular, gene transcript abundance in growth-related pathways, like ribosome biosynthesis and oxidative phosphorylation, varied significantly over the 2-d sampling period in these populations. Notably, we did not observe gene expression patterns that would suggest transition into the stationary phase over the 2-d sampling period. In total, these data suggest that frequent periods of metabolic acceleration and deceleration, even over the time span of only one doubling, may be a common modality in heterotrophic marine picoplankton species in situ.

The kinetics and regularity as well as quantitative and qualitative attributes of microbial response dynamics in situ have implications beyond biogeochemical considerations. Short time-scale microbe–environment and microbe–microbe interactions ultimately give rise to microbial population variation, functional variability, microbial community succession, and large-scale taxonomic shifts over days, weeks, and months and across seasons. A better understanding of short time-scale ecological microbial community processes should therefore provide a new perspective on longer-term community assembly, structure, and functional patterns. Future studies using the approaches we describe here have potential to yield deeper insight into microbial environmental interactions and their ecological consequences in a dynamic and constantly changing environment.

**Methods**

**Sample Collection.** Seawater samples (1 L) were collected along the central California coast in September of 2010 using an Environmental Sample Processor (ESP) (6, 10) suspended beneath a free-drifting surface float at 23-m depth. Microbes in the 0.22- to 5-μm size fraction were collected and preserved as previously described (6) but with a reduced incubation time in RNALater (Ambion) of 2 min per wash, which yields RNA of similar integrity. The instrument was recovered on September 19, 2010, and sample filters were moved to individual vials for long-term storage at −80 °C within 36 h.

**Library Preparation and Sequencing.** Approximately one-half of each filter was used for extraction of total community RNA and subtractive hybridization of rRNAs as previously described (28). Synthesis of antisense rRNA probes used DNA extracted from 5.8- to 7.1-L seawater samples collected using a rosette sampler at 23-m depth near the ESP at 10:00 AM on September 15, 17, and 19. Dilutions from these stocks were pooled and used for cDNA synthesis of bacterial, archaeal, and eukaryotic large- and small-subunit RNA probes. Approximately 150 ng total community RNA were hybridized with 300 ng each bacterial, 100 ng each archaeal, and 150 ng each eukaryotic small- and large-subunit probes. Probe removal used two successive 5-min incubations with 75 μL washed Streptavidin beads (NEB) in a final volume of 50 μL. Purified and concentrated message was linearly amplified and converted to cDNA as described previously (2).

A GS FLX Titanium system (Roche) was used to sequence cDNA. Library preparation followed the Titanium Rapid Library Preparation protocol. To improve the retention of smaller cDNA molecules, adaptor-ligated libraries were not diluted before size selection with AMPure XP beads. Libraries were quantified using the Titanium Slingshot kit (Fluidigm) and added to emulsion PCR reactions at 0.1 molecules per bead. Sequencing and quality control followed the manufacturer’s recommendations.

**Sequence Analysis and Annotation.** Our analytical pipeline for sequence annotation is summarized in SI Appendix, Fig. S3. Metatranscriptomic sequence libraries were screened for rRNA-derived transcripts and duplicates as previously described (28). Putative coding sequences with bit scores ≥ 50 were initially identified by BLASTX against the NCBI nonredundant peptide database as downloaded on May 31, 2010. After this initial analysis, additional reference sequences became available for SAR86 cluster *Gammaproteobacteria* and MGII *Euryarchaeaa*. All unique nonrRNA sequences were again compared by BLASTX with these newly released genome sequences, retaining those sequences with bit scores ≥50 that were greater than or equal to their best match in the previous NCBI database search (SI Appendix, Fig. S3). Sequence classification and annotation used the highest-scoring database match followed by the highest-scoring homologous match (with *Pelagibacter* and *HIMB114*, which we included within the *Pelagibacter*). For sequences matching equally well to multiple genes within the database, all matches were required to fall within the Chlorophyta for assignment to *Ostreococcus*, the Cyanobacteria for *Synechococcus*, and the SAR11 cluster for *Pelagibacter*. All top-scoring matches were required to fall within the SAR86 cluster or the MGII *Euryarchaeaa* for assignment to those taxon bins. Sequences were mapped to a single reference gene for annotation purposes, with preference given to references that were abundant in the dataset and references derived from sequenced genomes. Data files containing all taxon-specific transcript sequences for *Ostreococcus*, *Synechococcus*, *Pelagibacter*, SAR86, and Group II *Euryarchaeaa* are available from the authors on request.

Within each major taxonomic bin, sequence counts for genes present in multiple reference genomes were compiled to generate ortholog cluster-based transcript abundances. This approach was implemented to avoid artificial division of transcript pools from environmental organisms among multiple imperfectly matched reference sequences. Pairwise reciprocal best BLAST hits between translated coding sequences of reference genomes were compiled to generate ortholog cluster assignments. Identification of shared genes in *Ostreococcus* used the previously described e-value–based significance cutoff of 10−5 (29), whereas *Synechococcus*, *Pelagibacter*, and MGII comparisons used an e-value cutoff of 10−3 and required 30% alignment identity over 80% of the longer sequence; SAR86 comparisons used the 10−5 e-value and 30% identity cutoffs, but (because of the highly fragmented nature of the SAR86 assemblies) only 50% of the longer sequence was required to align. Func-
tional annotation of ortholog clusters used the Kyoto Encyclopedia of Genes and Genomes (KEGG) (30) annotations where available. Key species lacking curated annotations were analyzed using the KEGG automated annotation pipeline (31). In some cases, metatranscriptomic sequences were mapped to reference genes that were not derived from sequenced genomes (i.e., environmental clones). Where possible, these references were assigned to ortholog clusters based on single-directional peptide BLAST (significance cutoffs as above). Full lists of ortholog cluster membership, annotation, and results of statistical analyses are available in Datasets 51, 52, and 53.

Functional Clustering of Transcriptional Profiles. Cluster-based analyses were used to examine global patterns of transcription within and across taxon bins. Because many of these analyses assume normally distributed data, a variance-stabilizing transformation (32) was applied before analysis:

\[
x = \sin^{-1} \left( \frac{c}{N} \right) \quad \text{if } c > 0
\]

\[
x = \frac{1}{2} \sin^{-1} \left( \frac{N}{c} \right) \quad \text{if } c = 0
\]

\[
x = \frac{1}{2} \left( \sin^{-1} \left( \frac{2c}{N^2} \right) - \pi \right) \quad \text{if } c < 0
\]

where \(c\) is the count of each transcript and \(N\) is the library size at each time point. Mantel tests, principal components analyses, and Procrustes tests used variance-stabilized transcript abundances and were carried out using the vegan software package (33).

The geneARMA (16) package was used to identify global patterns of coregulation among Ostreococcus, Synechococcus, Pelagibacter SAR86, and MGI1 transcripts. This algorithm performs model-based soft clustering of transcriptional profiles using an autoregressive moving average model, ARMA(\(p,q\)), for the longitudinal covariance structure and Fourier series functions to model gene expression patterns. Data were filtered before analysis such that the maximum count for a row (ortholog time series) was greater than 5 and the sum of the row was greater than 10. For each dataset, the algorithm was run multiple times (100–400 iterations) from random initializations for \(p = 1–2\) autocorrelation terms, \(q = 0–1\) moving average terms, \(K = 1–3\) Fourier function terms, and \(J = 1–40\) clusters. The iteration possessing the highest likelihood for each parameter combination was used for final inference, and the Akaike Information Criterion was used to assess model complexity. The optimal model configuration, as identified by Akaike information criterion, for all five datasets was an ARMA(1,1) covariance structure and a three-term Fourier series mean function, and it included 39 clusters for Ostreococcus, 25 clusters for Synechococcus, 13 clusters for Pelagibacter, 7 clusters for SAR86, and 9 clusters for MGI1 (SI Appendix, Fig. S12).

Regression Tests for Count Data. Gene-by-gene tests to identify transcripts exhibiting sinusoidal periodicity or covariance with pathway-level functions used Poisson log-linear regression as implemented in the R software package (34). Library size offsets were based on the total number of transcripts assigned to a given taxon at each time point. For periodicity tests, the sinusoidal function \(x = A \cos(\omega t + \phi)\), where \(A\) represents the amplitude, \(\omega\) is the phase, and \(t\) is the midpoint of the sampling time in hours, was reduced to the linear equation \(x = A \cos(\omega t) + \sin(\omega t)\), where \(\alpha = A \cos \omega\) and \(\beta = A \sin \omega\). The significance of each model fit was assessed using both a \(\chi^2\) test (as implemented in the anova.glm function) and a permutation test. Permutation \(P\) values were calculated as the fraction of randomized datasets with a model fit (evaluated using the difference between the null and residual deviance) as good or better than model fits of the actual experimental data. To optimize computational resources, permutations continued until at least 10 randomized datasets with likelihood ratios equal to or exceeding the observed data had been identified (500–50,000 permutations). False discovery rate-corrected (35) \(P\) values of at least 0.1 from both tests were required for a relationship to be considered significant.

ACKNOWLEDGMENTS. This manuscript is dedicated to the memory of Carl R. Woese, who profoundly and forever changed our understanding of the evolution and nature of life on Earth. We thank the Environmental Sample Processor team for their help in planning and implementing this experiment, especially Gene Massion, Brent Roman, Scott Jensen, Roman Marin III, Chris Proctor, and Chad Birch. We also thank the engineers and technicians at Monterey Bay Aquarium Research Institute (MBARI) for their help and dedication to instrument development, and the crew of the RV Western Flyer for their support and expertise during field operations. This work is a contribution to the Center for Microbial Ecology: Research and Education (C-MORE). Development and application of the Environmental Sample Processor has been supported by National Science Foundation Grant OCE-0314222 (to C.A.S.), National Aeronautics and Space Administration and NASA’s Johnson Astrobiology Grants NN06G854G and NNG05G08G (to C.A.S.), the Gordon and Betty Moore Foundation (C.A.S.), and the David and Lucille Packard Foundation. This work was supported by National Science Foundation Science and Technology Center Award EF0242599 (to C.A.S. and E.F.D.), a grant from the Gordon and Betty Moore Foundation (to E.F.D.), and a gift from the Agouron Institute (to E.F.D.).


