Microbial community transcriptional networks are conserved in three domains at ocean basin scales

The MIT Faculty has made this article openly available. Please share how this access benefits you. Your story matters.
Microbial community transcriptional networks are conserved in three domains at ocean basin scales

Frank O. Aylwarda, John M. Eppleyb, Jason M. Smithb, Francisco P. Chavezb, Christopher A. Scholinb, and Edward F. DeLongabc1

*Daniel K. Inouye Center for Microbial Oceanography Research and Education (CA-MORE), University of Hawaii at Manoa, Honolulu, HI 96822; aMonterey Bay Aquarium Research Institute, Moss Landing, CA 95039; bDepartment of Civil and Environmental Engineering, Massachusetts Institute of Technology, Cambridge, MA 02139; and cDepartment of Biological Engineering, Massachusetts Institute of Technology, Cambridge, MA 02139

Contributed by Edward F. DeLong, February 13, 2015 (sent for review December 17, 2014; reviewed by Jeffrey I. Gordon, Curtis Huttenhower, and Thomas Schmidt)

Planktonic microbial communities in the ocean are typically dominated by several cosmopolitan clades of Bacteria, Archaea, and Eukarya characterized by their ribosomal RNA gene phylogenies and genomic features. Although the environments these communities inhabit range from coastal to open ocean waters, how the biological dynamics vary between such disparate habitats is not well known. To gain insight into the differential activities of microbial populations inhabiting different oceanic provinces we compared the daily metatranscriptome profiles of related microbial populations inhabiting surface waters of both a coastal California upwelling region (CC) as well as the oligotrophic North Pacific Subtropical Gyre (NPSG). Transcriptional networks revealed that the dominant photoautotrophic microbes in each environment (Ostreococcus in CC, Prochlorococcus in NPSG) were central determinants of overall community transcriptive dynamics. Furthermore, heterotrophic bacterial clades common to both ecosystems (SAR11, SAR116, SAR86, SAR406, and Roseobacter) displayed conserved, genome-wide inter- and intratranscriptional patterns and diel cycles. Populations of SAR11 and SAR86 clades in particular exhibited tightly coordinated transcriptional patterns in both coastal and pelagic ecosystems, suggesting that specific biological interactions between these groups are widespread in nature. Our results identify common diurnally oscillating behaviors among diverse planktonic microbial species regardless of habitat, suggesting that highly conserved temporally phased biotic interactions are ubiquitous among planktonic microbial communities worldwide.

Significance

Microbes drive biogeochemical cycles across the globe, collectively playing a central role in shaping the biosphere. Despite their immense importance, the in situ activities of communities of microbes, in particular uncultivated lineages of "microbial dark matter," remain poorly elucidated. In this study, we report that common temporal and ecological dynamics underpin disparate marine microbial communities, providing the first evidence that trans-Pacific diurnal transcriptional patterns in these communities may regulate ecological and biogeochemical processes across the ocean. In total, our findings indicate a remarkable similarity in the timing of community-wide activity in the ocean, and suggest that global patterns of a variety of biogeochemical transformations may be temporally predictable and governed by structured ecological determinants.


Reviewers: J.J.G., Washington University School of Medicine in St. Louis, CH.H., Harvard School of Public Health; and T.S., University of Michigan. The authors declare no conflict of interest.

Freely available online through the PNAS open access option.

Data deposition: The sequence reported in this paper has been deposited in the NCBI Sequence Read Archive database (accession no. SRP050269).

1To whom correspondence should be addressed. Email: edelong@hawaii.edu.

This article contains supporting information online at www.pnas.org/cgi/doi/10.1073/pnas.1502883112/DCSupplemental.

www.pnas.org/cgi/doi/10.1073/pnas.1502883112
Results and Discussion

As expected, different phototrophic primary producers dominated the coastal versus open ocean metatranscriptomes (Ostreococcus in CC and Prochlorococcus in NPSG) (15–17 (SI Appendix, Fig. S1 A and B and Dataset S3)). In contrast, many dominant heterotrophic bacterioplankton clades were common to community transcripts in both habitats, including the SAR11, SAR86, SAR116, SAR406, and Roseobacter groups (SI Appendix, Fig. S1 A and B). We also identified populations of Flavobacteria, Marine Group II (MGII) Euryarchaeota, SAR92 and ARCTIC96-BD19 in the new coastal dataset, and a previously reported SAR324 population in the pelagic samples (Table 1 and SI Appendix, Fig. S1 A and B) (9). Mapping of reads onto reference genomes (95% identity, minimum 150 bp match length) indicated the overall phylogenetic profiles for the coastal and pelagic metatranscriptomes were largely non-overlapping (SI Appendix, Fig. S2), suggesting that although common heterotrophic clades were shared between the two habitats, their population-level genotypes were distinct. During the California coast drift, the relative abundance of reads assigned to Ostreococcus and MGII Euryarchaeota declined, whereas those assigned to Flavobacteria, SAR92, and ARCTIC96-BD19 increased. These changes were concomitant with increases in salinity and fluctuations in temperature (SI Appendix, Fig. S1C), potentially indicating movement of the ESP through distinct water masses. Similar albeit less severe variability in temperature and salinity was observed in the NPSG drift, although transcript abundance across phylogenetic groups was more stable in this dataset (SI Appendix, Fig. S1 B and D).

We analyzed the temporal dynamics of reads mapping to 8,844 and 7,276 orthologous protein clusters (henceforth referred to as “transcripts”) with high abundance in the coastal and pelagic metatranscriptomes, respectively, and compared the daily transcriptional patterns within and between different phylogenetic groups using a weighted transcriptomic network approach as implemented by the WGCNA package (18, 19) (Methods). Results of these analyses suggested that the majority of transcripts in both the coastal and pelagic transcriptomic networks clustered around the transcripts derived from the dominant phototrophic members of each community (Fig. 1 A and D). Moreover, Ostreococcus (CC) and Prochlorococcus (NPSG) transcripts were found to consistently have the highest connectivity (a.k.a. degree centrality) of all nodes in the networks (Fig. 2 A and Dataset S1), likely reflecting the central role of phototrophic community members in determining the overall community transcriptional profiles in both environments. Analysis of sets of transcripts with similar temporal profiles (termed modules; colored groupings in Fig. 1 B and E) revealed distinct patterns of 24-h periodic expression (Fig. 1 C and F), consistent with the underlying structuring of the transcriptomic networks around the well defined diel cycles of the phototrophic members.

All transcriptional modules identified within weighted networks (34 in CC, 18 in NPSG) included representative transcripts from numerous phylogenetic groups, indicating that similar, highly regulated transcriptional coordination was shared between microbial populations (SI Appendix, Fig. S3). Moreover, each module contained a broad representation of functional annotations both within and between taxa, reflecting coordinated regulation across metabolic pathways in different phylogenetic groups (SI Appendix, Figs. S4 and S5). Transcripts from different taxa involved in cofactor and vitamin metabolism and DNA processing often clustered in the same modules, as did those involved in oxidative phosphorylation and protein and RNA processing (SI Appendix, Figs. S4 and S5). Analysis of the functional annotation of transcripts across phylogenetic groups revealed that amino acid metabolism, DNA processing, translation, transporters, and vitamin metabolism consistently had the highest connectivity, indicating that these are potentially among the most tightly regulated metabolic processes (Fig. 2B). This pattern could be the result of intense competition for resources in both coastal and pelagic surface waters, which would render nutrient availability and the timing of growth and DNA replication as key determinants of the temporal dynamics in these planktonic assemblages. The abundance and connectivity of transcripts appeared to be associated in the pelagic but not coastal datasets (SI Appendix, Fig. S6), possibly reflecting a propensity of central metabolic genes critical for cellular processes to be highly expressed, consistent with previous findings (20). These patterns may also be due in part to a tendency for connectivity patterns to be more easily identified in highly expressed transcripts, although a lack of connectivity-abundance association in the coastal dataset suggests this is not an overriding factor. Together with the high connectivity of transcripts belonging to phototrophic groups, these results suggest that transcriptomic networks of both coastal and pelagic environments are largely structured by a combination of the diel cycles of the phototrophic members together with resource limitation common to both habitats.

In addition to shared, intrahabitat community-wide transcriptomic dynamics, similar transcriptional patterns were also observed between the heterotrophic bacterioplankton of coastal and pelagic environments. Using the transcriptomic networks as a foundation, we calculated the correlation of nodes to identify co-regulated genes as key determinants of the temporal dynamics in these ecosystems. This was particularly apparent for the SAR11 and SAR86 populations, which in both environments displayed tightly linked transcriptional profiles across a variety of functions including translation, transcription, and oxidative phosphorylation (Fig. 3). A similar but less apparent relationship was also found between the SAR116 and Roseobacter clades in both habitats (Fig. 3). In contrast, although SAR116 transcription patterns appear closely linked to Ostreococcus in the California coast, a similar relationship between SAR116 and Prochlorococcus was not observed in pelagic samples, suggesting

<table>
<thead>
<tr>
<th>Parameter</th>
<th>California coast</th>
<th>North Pacific Subtropical Gyre</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time points</td>
<td>35</td>
<td>30</td>
</tr>
<tr>
<td>Approx. sampling frequency</td>
<td>4 h</td>
<td>2 h</td>
</tr>
<tr>
<td>Duration of drift</td>
<td>9/12–17, 2012 (5 1/2 d)</td>
<td>9/7–11, 2011 (2 1/2 d)</td>
</tr>
<tr>
<td>No. of transcripts*</td>
<td>623 (50)</td>
<td>—</td>
</tr>
<tr>
<td>Flavobacteria</td>
<td>376 (1)</td>
<td>—</td>
</tr>
<tr>
<td>Gil Euryarchaeota</td>
<td>689 (34)</td>
<td>—</td>
</tr>
<tr>
<td>Roseobacter</td>
<td>867 (99)</td>
<td>542 (165)</td>
</tr>
<tr>
<td>Ostreococcus</td>
<td>1,602 (781)</td>
<td>—</td>
</tr>
<tr>
<td>Prochlorococcus</td>
<td>—</td>
<td>2,052 (1,464)</td>
</tr>
<tr>
<td>SAR11</td>
<td>1,300 (296)</td>
<td>1,223 (270)</td>
</tr>
<tr>
<td>SAR86</td>
<td>1,400 (31)</td>
<td>795 (39)</td>
</tr>
<tr>
<td>SAR92</td>
<td>404 (2)</td>
<td>—</td>
</tr>
<tr>
<td>SAR116</td>
<td>969 (146)</td>
<td>1,661 (146)</td>
</tr>
<tr>
<td>SAR234</td>
<td>513 (28)</td>
<td>—</td>
</tr>
<tr>
<td>SAR406</td>
<td>615 (0)</td>
<td>490 (17)</td>
</tr>
</tbody>
</table>

*Total number of transcripts (those with significant 24-h periodicity).
that interactions between this group and different photoautotrophic primary producers may not be highly conserved.

We identified numerous transcripts from heterotrophic groups that displayed significant 24-h periodicity in both coastal and pelagic metatranscriptome time series, consistent with a recent report for the NPSG dataset (9) (Table 1). The diel signal in heterotrophic groups in both environments was most prevalent in transcripts associated with translation and oxidative phosphorylation (Fig. 1B), which are useful indicators of overall metabolic activity (21). In both the coastal and pelagic habitat, the peak abundance of photoautotroph carbon fixation and photosynthesis transcripts was followed by a tight “cascade” of translation and oxidative phosphorylation transcript maxima from heterotrophic groups (Fig. 1B). This pattern of daily transcript maxima that proceeded temporally from *Roseobacter* to SAR116 to SAR11 and SAR86 was strikingly similar between coastal and pelagic environments, suggesting conservation of temporal associations in community-wide diel cascades among these phylogenetic groups across the Pacific Ocean basin. The pattern was less compressed and occurred earlier in the day in pelagic populations, probably due to the earlier daily light influx (Photosynthetically Available Radiation, PAR; Fig. 3A) that signals the start of photosynthesis in this habitat. Although periodic patterns in lower abundance groups were more difficult to discern, their general timing of translation and oxidative phosphorylation transcript maxima were similar to the other more abundant heterotrophic groups. Overall, the timing of these events suggested that regardless of taxon, the activities of abundant photoautotrophs in both coastal and pelagic habitats “set” the whole-community diel cycle.

We postulate that for heterotrophs this pattern of translation/oxidative phosphorylation transcript maxima is initiated and sustained by the progressive release of key nutrients from primary producers, followed by successive, temporally ordered metabolic conversions mediated by different heterotrophic community members throughout the day. For heterotrophic
groups, the same transcriptional modules that contained the most translation and oxidative phosphorylation transcripts (9 and 15 in CC, 3 in NPSG) also contained a number of transcripts belonging to predicted sugar and amino acid transporters, peptidases, and genes involved in amino acid metabolism and the citric acid cycle (SI Appendix, Figs. S4 and S5 and Dataset S1),

Fig. 2. Transcripts from both the California coast and NPSG metatranscriptomes belonging to photoautotrophic community members (A) and those involved in key metabolic processes such as amino acid metabolism, DNA processing, and translation (B) have the highest connectivity (degree centrality) in transcriptional networks. The scaled connectivity (degree centrality) of transcripts is given on the y-axes, with transcripts ordered highest to lowest based on taxonomic assignment and KEGG annotation. Met: Metabolism; Phosph.: Phosphorylation.

Fig. 3. Patterns of interspecies transcriptional coordination are conserved across coastal and pelagic planktonic microbial communities. Correlated and anticorrelated functional pathways in the photoautotrophic and shared heterotrophic groups are shown for the California coast (CC, Upper Left) and North Pacific Subtropical Gyre (NPSG, Lower Right) metatranscriptomes. Pathway correlations were calculated from scaled adjacency matrices by averaging across the correlations of the cognate transcripts (Methods). Pathways: 1: Transporters; 2: Translation; 3: Transcription; 4: Protein/RNA processing; 5: Oxidative phosphorylation; 6: Cofactor and vitamin metabolism; 7: Citrate cycle; 8: Amino acid metabolism; 9: Chemotaxis and flagellar assembly; 10: Central carbon metabolism; 11: Glycolysis; 12: DNA and chromosome processing.
suggesting that an increase in the availability of certain compounds in marine dissolved organic matter (DOM) may spur their transcriptional upsurges. Supporting this hypothesis, distinct transcriptional responses to elevated DOM have been shown to occur in heterotrophic bacterioplankton studied in microcosm experiments (22). Given that Ostreococcus and Prochlorococcus likely produce distinct suites of dissolved organic matter (23), this would suggest that the temporal succession of microbial activity may exist regardless of the specific compounds involved.

Metabolic patterns in heterotrophic groups could also be inferred from other pathways. In particular, reads mapping to the glyoxylate shunt enzymes isocitrate lyase and malate synthase were highly abundant in both metatranscriptomes (SI Appendix, Fig. S7), consistent with findings that this pathway plays a key role in carbon provisioning in bacterioplankton (24). Overall, transcripts mapping to ribosomal proteins, oxidative phosphorylation enzymes, and diverse transporters were relatively more abundant in coastal than pelagic SAR11, SAR116, SAR86, SAR406, and Roseobacter populations (Mann–Whitney U Test, $P < 0.005$) (SI Appendix, Fig. S8 and Dataset S2), likely reflecting the higher nutrient availabilities in the coastal system. In contrast, transcripts mapping to proteins involved in light stress attenuation (photo-lyases), free radical detoxification (catalase, peroxiredoxins, thioredoxins), photoactive damage of coenzymes (biotin sulfoxide reductase), and heat shock in all of these same groups were more highly expressed in pelagic than coastal populations (SI Appendix, Fig. S8). These results are consistent with the higher solar radiation and temperatures that typify surface waters of the NPSG, supporting the recently recognized importance of heterotrophic free radical detoxification processes in this ecosystem (25, 26). Transcripts involved in phosphonate uptake were also enriched in pelagic versus coastal SAR11 and SAR116 populations (Dataset S2), reflecting their importance as a source of phosphate in oligotrophic environments (27, 28).

Significant progress has been made over the past decades in characterizing the phylogenetic diversity of eukaryotic, bacterial and archaeal groups inhabiting the world’s oceans (3, 29, 30), but the in situ metabolic activities and interactions of these groups that combine to create cohesive, resilient community structure...
and ecosystem-level processes remain largely undetermined. The extraordinary similarity of transcriptional temporal dynamics and coordination that is shared between microbial populations of disparate marine ecosystems as shown here illustrates the importance of short-term temporal interactions in defining the ecological niches of microbial plankton. Moreover, the rapid turnover of nutrients in the upper ocean (14, 31) together with the observed seasonality in whole-community diel cycles among all three domains of life in these disparate habitats implies that biogeochemical fluxes mediated by microbial community metabolism are diurnally phased across entire ocean basins. Identifying the precise mechanisms and metabolic fluxes that regulate the specific microbial interactions giving rise to whole-community metabolism is a major challenge for microbial oceanography in the future. Our findings underscore the importance of gaining deeper mechanistic perspective into the tightly coordinated ecosystem-level dynamics that orchestrate daily biogeochemical fluxes, and how they may respond in the face of rapid anthropogenic environmental change.

Materials and Methods

ESP Deployment and Sample Collection. For our coastal metatranscriptomes, samples were collected by an Environmental Sample Processor (ESP) (12) suspended at a depth of 10 m from September 12–17, 2012 off the coast of California near Monterey (initial deployment: −122.7°W, 36.4°N). As previously described, planktonic microbial assemblages were collected in the 0.22-μm size range at 4-h intervals and preserved aboard the ESP (10, 12). Filter samples collected by the ESPs were stored at −80 °C within 24 h of their retrieval. RNA was extracted, CDNA was generated, and Illumina sequencing (32) was performed as described (33–35). Details regarding sequence processing and classification can be found in the Supplementary Appendix. Sequences generated in this study have been deposited in the NCBI Sequence Read Archive under accession number SRP050269. Collection and processing of the NPSG samples has been described (9).

Transcriptomic Network Analyses. For both the coastal and pelagic metatranscriptomes reads were mapped to ortholog clusters of proteins constructed from the phylogenetic groups of interest, and a weighted transcriptomic network approach using the R packages WGCNA and igraph was used to analyze the resulting count tables. Details regarding the soft thresholds used in network construction, the calculation of adjacency matrices, and the partitioning of transcripts into modules can be found in SI Appendix.

Acknowledgments. We thank B. Barone, D. Mende, T. Nielsen, T. Palden, A. Rosano, S. Wilson, and members of the E.F.D. laboratory for helpful comments, J. Robidart for collecting seawater samples for RNA subtraction-probe construction, and E. Ottesen for sharing R code. This work was supported by grants from the Gordon and Betty Moore Foundation (to E.F.D., 3777), National Science Foundation Grant EF0424599 (to E.F.D.), and the Simons Foundation grant Simons Collaboration on Ocean Processes and Ecology (SCOPE) (to E.F.D.). Development and application of the Environmental Sample Processor has been supported by National Science Foundation Grant OCE-0314222, National Aeronautics and Space Administration Astrobiology Grants NNG06GB34G and NNX09AB78G, the Gordon and Betty Moore Foundation (to C.A.S., 2728), and the David and Lucille Packard Foundation (to C.A.S., 901204, 901205). The cruise in the California Current was part of the MBARI CANON initiative and supported by the David and Lucille Packard Foundation (to F.P.C.; 901009). This work is a contribution of the Center for Microbial Oceanography: Research and Education (C-MORE) and the Simons Collaboration on Ocean Processes and Ecology (SCOPE).

10. Karl DM, Church MJ (2014) Microbial oceanography in the future. Our findings underscore the importance of gaining deeper mechanistic perspective into the tightly coordinated ecosystem-level dynamics that orchestrate daily biogeochemical fluxes, and how they may respond in the face of rapid anthropogenic environmental change.

Materials and Methods

ESP Deployment and Sample Collection. For our coastal metatranscriptomes, samples were collected by an Environmental Sample Processor (ESP) (12) suspended at a depth of 10 m from September 12–17, 2012 off the coast of California near Monterey (initial deployment: −122.7°W, 36.4°N). As previously described, planktonic microbial assemblages were collected in the 0.22-μm size range at 4-h intervals and preserved aboard the ESP (10, 12). Filter samples collected by the ESPs were stored at −80 °C within 24 h of their retrieval. RNA was extracted, CDNA was generated, and Illumina sequencing (32) was performed as described (33–35). Details regarding sequence processing and classification can be found in the Supplementary Appendix. Sequences generated in this study have been deposited in the NCBI Sequence Read Archive under accession number SRP050269. Collection and processing of the NPSG samples has been described (9).

Transcriptomic Network Analyses. For both the coastal and pelagic metatranscriptomes reads were mapped to ortholog clusters of proteins constructed from the phylogenetic groups of interest, and a weighted transcriptomic network approach using the R packages WGCNA and igraph was used to analyze the resulting count tables. Details regarding the soft thresholds used in network construction, the calculation of adjacency matrices, and the partitioning of transcripts into modules can be found in SI Appendix.

Acknowledgments. We thank B. Barone, D. Mende, T. Nielsen, T. Palden, A. Rosano, S. Wilson, and members of the E.F.D. laboratory for helpful comments, J. Robidart for collecting seawater samples for RNA subtraction-probe construction, and E. Ottesen for sharing R code. This work was supported by grants from the Gordon and Betty Moore Foundation (to E.F.D., 3777), National Science Foundation Grant EF0424599 (to E.F.D.), and the Simons Foundation grant Simons Collaboration on Ocean Processes and Ecology (SCOPE) (to E.F.D.). Development and application of the Environmental Sample Processor has been supported by National Science Foundation Grant OCE-0314222, National Aeronautics and Space Administration Astrobiology Grants NNG06GB34G and NNX09AB78G, the Gordon and Betty Moore Foundation (to C.A.S., 2728), and the David and Lucille Packard Foundation (to C.A.S., 901204, 901205). The cruise in the California Current was part of the MBARI CANON initiative and supported by the David and Lucille Packard Foundation (to F.P.C.; 901009). This work is a contribution of the Center for Microbial Oceanography: Research and Education (C-MORE) and the Simons Collaboration on Ocean Processes and Ecology (SCOPE).