Microbial Metatranscriptomics: towards Understanding Microbial Gene Expression and Regulation in Natural Habitats

By

Yanmei Shi

B.S. Environmental Science
Nanjing University, 2003

M.S. Marine, Estuarine, and Environmental Sciences
University of Maryland, College Park, 2005

Submitted to the Department of Civil and Environmental Engineering
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy
in Civil and Environmental Engineering
at the

Massachusetts Institute of Technology

February 2011

Author: _____
Department of Civil and Environmental Engineering
January 3, 2011

Certified by: _____
Edward F. DeLong
Morton and Claire Goulder Professor in Environmental Systems
Professor of Civil and Environmental Engineering & Biological Engineering
Thesis Supervisor

Accepted by: _____
Heidi Nepf
Chair, Departmental Committee for Graduate Students

Microbial Metatranscriptomics: towards Understanding Microbial Gene Expression and Regulation in Natural Habitats

by

Yanmei Shi

Submitted to the Department of Civil and Environmental Engineering in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the field of Environmental Biology

Abstract

Metagenomic research has paved the way for a comprehensive understanding of the microbial gene parts list in nature, but a full understanding of microbial gene expression, regulation, and ecology remains a challenge. In this thesis, I present the methodological foundations and applications of deep sequencing-based metatranscriptomics, for profiling community transcriptomes on spatial and temporal scales. Several findings and relevant hypotheses have emerged from this work. I show that transcripts of house-keeping genes necessary for the maintenance of basic cellular machinery are abundant and readily detectable. Habitat-specific transcripts are also discernible when comparing community transcriptomes along distinct geochemical conditions. Normalization of detected transcripts to their corresponding gene abundance suggests that numerically less abundant microorganisms may nevertheless contribute actively to ecologically relevant processes. Along the same lines, it is a recurrent observation that many transcripts are of unknown function or phylogenetic origin, and have not been detected in genomic/metagenomic data sets. These novel sequences may be derived from less abundant species or variable genomic regions that are not represented in sequenced genomes. Furthermore, I applied metatranscriptomics in a microcosm experiment, where a deep water mixing event was simulated and community transcriptomes were monitored over the course of 27 hours. Relative to the control, the treatment sample showed signals of stimulated photosynthesis and carbon fixation by phytoplankton cells, enhanced chemotactic, motility, and growth responses of heterotrophic bacteria, as well as possibly altered phage-host interactions. Such experimental metatranscriptomic studies are well suited to reveal how microorganisms respond during the early stages of environmental perturbations. Finally, I show that metatranscriptomic data sets contain a wealth of highly expressed small RNAs (sRNAs), transcripts that are not translated to proteins but instead function as regulators. I propose a bioinformatics pipeline for identifying these sRNA elements, characterizing their structures and genomic contexts, and predicting possible regulatory targets. The extraordinary abundance of some of the identified sRNAs raises questions about their ecological function, which warrants further biochemical and genetic studies. Overall, this work has extended our knowledge of functional potentials and *in situ* gene expression of natural microbial communities.

Thesis Supervisor: Edward F. DeLong
Title: Morton and Claire Goulder Professor in Environmental Systems
Professor of Civil and Environmental Engineering & Biological Engineering

Acknowledgements

I am sincerely grateful to all the people who have helped and supported me all the way along my Ph.D. endeavor; without their companionship the journey would have been much more difficult and less fun.

First I deeply appreciate the guidance, support, and academic freedom my advisor Edward DeLong has provided me all through my thesis work, even when I was sidetracked by scientific questions and topics that were not the focus of the lab. His trust on my academic capability has been a constant motivation for me. I also want to thank my committee members, Martin Polz, Eric Alm, Mak Saito, and (previously) Dianne Newman, for their support and advice on my research and career.

I am very fortunate to have joined a group of caring, fun, and hardworking colleagues, including current and past members in the DeLong lab and in Parsons. Kostas Konstantinidis taught me about bioinformatics, which I knew nothing about when I started embarking on UNIX and perl programming. Gene Tyson, during his two years at MIT, was an amazing collaborator and mentor, to whom I bear great gratitude. Chon Martinez is an indispensable person in the lab, and I thank her for all the advice she gave me, not only on PhD work but also on career path. Virginia Rich has been a dear friend and collaborator, with whom I share warm memories both personally and professionally. John Eppley has been the go-to person whenever I have computational problems. Laure-Anne Ventouras has been a dear friend to me, in and outside of the lab, and I am glad to have shared so many good times with her. There are so many other colleagues I want to thank — Jay McCarren, Julie Maresca, Frank Stewart, Elizabeth Ottesen, Adrian Sharma, Rex Malmstrom, Jorge Frias-Lopez, Vinh Pham — just to name a few, for their enthusiasm and support.

I want to thank other friends, for sharing good food, movies, traveling, and joyful conversations. Juan Guan (an unfailing source of joy, humor and frequently sarcasm), Hong Xue and Su Xu (and their lovely son Yuran Xue), Sam Wilson and Pancho Bustos (such happy days in Hawaii!), Lu Gao (wish I knew you earlier), Taeko Minegishi (thank you for always being there), Jane Kim (my former roomie), Qiong Yang (we have a lot in common) — I thank you all for bringing color and diversity to my life!

My family has been a constant source of strength, comfort, and joy to me. I thank my dear parents, Tongming Shi and Meifang Ge, for their unconditional love. I also appreciate the love and support my in-laws give me. My husband, Bin Wang, you are my inspiration and motivation for longer than I can remember. Our dear precious daughter Chloe Zhiyue Wang, you don't know this yet, but you mean everything to me.

This work was supported by a grant from the Gordon and Betty Moore Foundation (EFD), the Office of Science (BER) U. S. Department of Energy (EFD), and NSF Science and Technology Center Award EF0424599. This work is a contribution of the Center for Microbial Oceanography: Research and Education (C-MORE).

Table of Contents

CHAPTER ONE

Introduction

Chapter 1: Introduction

Overview

Microorganisms represent major numerical and functional components in essentially every habitat on Earth. Microbial cells were estimated to contain roughly 10^{17}-10^{18}g, 10^{17}g, and 10^{16}g of C, N, and P, respectively (Whitman, Coleman & Wiebe, 1998); thus the growth and turnover of naturally occurring microorganisms represent a significant and active part of global nutrient cycling. In addition, an estimated 10^{30}-10^{31} bacterial and archaeal cells around the world are actively mediating essential ecological processes. Understanding their metabolic capabilities and activities are therefore fundamental to understanding the functioning of the Earth system.

Microorganisms in nature rarely live alone, but instead function as integrated units (communities) that interact with one another and with their surrounding environments. Over the past three decades, the use of molecular phylogenetic approaches has profoundly changed our view of microbial diversity, revealing a wealth of uncultivated microbial species (Curtis, Sloan & Scannell, 2002; Pace, 1997) or ecologically-coherent units (Acinas et al., 2004; Hunt et al., 2008). Metagenomic surveys (collection and analyses of community DNA without cultivation) further revealed an enormous and dynamic pool of microbial genes (metabolic capabilities) harbored by these microbial assemblages (DeLong et al., 2006; Rusch et al., 2007). Naturally, the next step is to understand how such genomic and metabolic diversity is expressed (or in other words, manifested at the community transcriptome level) on temporal and spatial scales. Numerous genome-wide expression studies have been performed with laboratory cultures under the settings of both natural science and medical researches (Ernst et al., 2005; Sharma et al., 2010; Toledo-Arana et al., 2009; Zinser et al., 2009). These studies have yielded invaluable information about gene expression organization and dynamics, facilitating the use of transcriptomes as indicators for cell physiology or diagnosis tools for diseases. Similarly, an important goal of studying microbial community transcriptomes in natural habitats is to be able to use them as probe and sensor to predict changes in microbial community dynamics during natural or anthropogenic environmental perturbations—such as seasonal changes or global climate change.

In this Chapter, I first present a brief introduction to the research context of community

transcriptomics (aka, metatranscriptomics), followed by an overview of the recent development and advancement of metatranscriptomic studies (Figure 1). Next, I highlight major findings revealed by metatranscriptomic surveys and defined experiments. Just like with any other scientific researches, findings and conclusions should be presented in the context of methodology (Figure 2), potential limitations, and space for future improvements. Finally, to put this thesis in context, I lay out and briefly describe the structure of the main body of the thesis.

Why Metatranscriptomics?

The advent of cultivation-independent metagenomic approach has provided apparently inexhaustible access to microbial diversity - both phylogenetically and functionally (Brazelton et al., 2010; DeLong et al., 2006; Tringe & Rubin, 2005; Turnbaugh et al., 2009; Warnecke et al., 2007). The healthy debate of the extent of such diversity is beyond the scope of this chapter, but it is generally agreed that little is known about the functional significance of the observed genes. What genes are being expressed by what organisms, to what extent, when, and where? These are important questions, the answers to which provide one step further in decoding microbial activities *in situ*.

Interest in understanding microbial gene expression *in situ* is not new, but the depth of our knowledge has been constrained by the available methods for observing it. Conventionally, reverse transcription quantitative PCR (RT-qPCR) was the main tool to detect and quantify transcripts in the environment. The use of RT-qPCR requires prior knowledge of sequences (including their variants) of targeted genes, in order to design primers and probes that allow detection of a range of orthologs. In addition, the technology setup is low-throughput regarding the number of targeted genes, most of which are involved in well studied pathways such as N/P metabolism and photosynthesis (Church, Wai, Karl & DeLong, 2010; Orchard, Webb & Dyhrman, 2009; Steunou et al., 2006). Inspired by the successful use of microarray technique in quantifying genome-wide expression (for hundreds to thousands of genes simultaneously) (Lindell et al., 2007; Sharma et al., 2010; Zinser et al., 2009), researchers have developed versions of environmental functional microarrays in efforts to overcome the gene number constraints. These microarrays harbor thousands to tens of thousands of probes either selected from sequenced genomes (Parro, Moreno-Paz & González-Toril, 2007) or randomly selected

from environmental cDNA clone libraries (McGrath et al., 2010). Nevertheless, microarray's technological limitations persist (Zhou & Thompson, 2002). These challenges include dependence on massively parallel nucleic acid hybridization, potential for cross-hybridization of highly related sequences, complex and often indirect quantification algorithms, and outputs as signal intensity but not nucleotide sequence identities. For all these reasons, attempts have been made to profile community transcripts in a non-targeted and sequence-based fashion. In 2005, Poretsky and colleagues generated a cDNA clone library by random priming of microbial community RNA collected from a hypersaline lake, and sequenced the library, although the scale was relatively limited (~ 400 clones) (Poretsky et al., 2005).

Next-generation sequencing techniques, such as pyrosequencing (Margulies et al., 2005), Illumina technology (formerly Solexa sequencing), and more recently Ion Torrent technology (Ion Torrent Systems, Inc., Guilford, CT), enable producing millions of sequence reads in a single run, and hence represent a fundamental leap towards large-scale, sequence-based profiling of community transcriptomes (Mardis, 2008). Since pyrosequencing (Margulies et al., 2005) was first used to assess community transcripts in soil samples (Leininger et al., 2006), this approach has gained a lot of popularity in the microbial ecology field, with dozens of peer-reviewed metatranscriptomics publications since 2007 (Figure 1). More than half of these published studies were focused on open ocean microbial assemblages, due in part to the relative ease in size-fractionating and collecting bacterioplankton biomass.

For metatranscriptomic methods based on next-gen sequencing, total RNA is extracted from a microbial community, processed as needed (such as rRNA subtraction, amplification), converted into cDNA, and sequenced without the need for cloning (Figure 2). Protocols are continuously evolving (Figure 1), as new methodological and technological improvements arise (He et al., 2010b; Stewart, Ottesen & DeLong, 2010; Wu, Gao, Zhang & Meldrum, 2010). Its application is also expanding, from environmental surveys (Frias-Lopez et al., 2008) to comparative studies (Poretsky et al., 2009), and to experiments with well-defined perturbations (McCarren et al., 2010; Vila-Costa et al., 2010).

Gene expression and regulation reflected at the community level: What have we learned from metatranscriptomics?

Deep sequencing of bacterial transcriptomes, especially those of bacteria with small genomes (Guell et al., 2009; Sharma et al., 2010), have altered our view of the extent and complexity of bacterial transcription and regulation. Early conventional views of bacterial gene expression painted a fairly straightforward picture of transcriptional principles such as operon structure, promotors, and protein transcriptional regulators. Now, a more complex picture is emerging: anti-sense transcripts, alternative transcripts, variable transcriptional start sites, and regulatory small noncoding RNAs (sRNAs), are all prevalent signals in the deep-sequenced transcriptomes. Extrapolating from these, we expect community transcriptomes to be highly complex and informative with respect to the range and diversity of modes and mechanisms associated with microbial gene expression.

Emerging signatures shared by metatranscriptomes from distinct geochemical habitats. A somewhat surprising finding thus far is the presence of common features across metatranscriptomic data sets, despite the distinct geochemical conditions of the sampling sites. These shared signatures, reflected at the community level, point to some universal patterns of bacterial and archaeal gene expression in nature. Assuringly, classical models of bacterial and archaeal gene expression such as operon structure are apparent in metatranscriptomic data ((Poretsky et al., 2009); Coleman, PhD thesis). Some studies also showed evidence on less-established models such as the correlation between GC content, codon usage, sequence conservation, and gene expression ((Poretsky et al., 2009); Stewart *et al*, in preparation).

Based on functional representation, metatranscriptome samples tend to cluster together to the exclusion to their corresponding metagenome samples (Chapter 3; Stewart *et al* 2010, Environmental Microbiology, in press). This appeared to be caused, at least in part, by the active expression of house-keeping genes necessary for the maintenance of basic cellular machinery. Additionally, many of the transcript sequences that are of unknown functions or phylogenetic affiliations have not been detected or only detected in very low abundance in public metagenomic data sets (Chapter 2; (Gilbert et al., 2008)), further contributing to the separation of metatranscriptome and metagenome samples. This being said, metatranscriptome samples among themselves often cluster by habitat or environmental condition similarity, highlighting the expression of genes that are habitat-specific and ecologically relevant.

The abundance distribution of transcripts often follows a steep curve (i.e., the most abundant transcripts can be orders of magnitude more abundant than the least abundant transcripts; Chapter 2; Chapter 3). In addition, the tail representing low-abundance transcripts is very long: more than 25% of genes with detected transcripts are represented by only 1 sequence read (Stewart *et al* 2010, Environmental Microbiology, in press; Coleman, PhD thesis); about 66-74% of sequences with putative taxonomic assignment belonged to the top two most abundant taxonomic groups (Chapter 3). This recurring observation underlines that the sequencing depth of metatranscriptomes is far from saturating. The most highly expressed genes include house-keeping genes (e.g., ribosomal proteins, translation elongation factors), genes involved in habitat-specific process (e.g., ammonia monooxygenase genes), genes with unknown functions (and sometimes from low-abundance microorganisms that are not captured in the corresponding metagenomes), and noncoding intergenic regions (small RNAs, see below). These findings have already and are likely to continue to spur future research into unknown aspects of microbial transcriptomes in nature (Brown & Hewson, 2010; Shi, Tyson & DeLong, 2009).

Expect the unexpected: a wealth of highly expressed novel small RNAs. The term of "transcriptome" was originally defined as the complement of mRNAs transcribed from a cell's genome (Abbott, 1999). It is not accurate in a number of ways, as accumulating research has revealed a diverse and complex array of RNAs in bacterial and archaeal transcriptomes, that includes mRNA, tRNA, rRNA, anti-sense transcripts, and a variety of noncoding transcripts (Guell et al., 2009; Sharma et al., 2010; Steglich et al., 2008). Nonetheless, the presence of very highly expressed novel small RNAs (sRNAs) in metatranscriptomic data sets is an unexpected finding (Shi et al., 2009; Weinberg, Perreault, Meyer & Breaker, 2009), that has been a recurrent observation in all metatranscriptomic data sets.

Rapid and efficient regulation of gene expression is critical to environmental sensing and response of microbes in a dynamically changing environment. In recent years, an increasing number of studies have demonstrated that small RNAs (sRNAs) play critical regulatory roles in bacteria and archaea (Gottesman, 2002; Storz & Haas, 2007; Waters & Storz, 2009). Microbial sRNAs are untranslated short transcripts that are generally transcribed from intergenic regions and typically range from 50-500 bp in length. In model microorganisms such as *Escherichia coli*, *Vibrio cholerae* and *Bacillus subtlis*, 10-100 sRNAs have been experimentally identified and

hundreds more have been bioinformatically predicted (Livny, Fogel, Davis & Waldor, 2005; Silvaggi, Perkins & Losick, 2006; Vogel et al., 2003). Microbial sRNAs show a dramatic regulatory versatility: they are involved in the regulation of diverse pathways including oxidative responses, carbon storage, iron homeostasis, quorum sensing, and photosynthesis (Duehring, Axmann, Hess & Wilde, 2006; Gottesman, 2004; Lenz et al., 2004; Mandin & Gottesman, 2010). Additionally, the mechanisms by which microbial sRNAs act are very diverse. Most sRNAs bind to untranslated regions (UTR) of target genes with specificity achieved by (often imperfect) base-pairing interactions, and consequently affect gene transcription, mRNA stability, and translation. However, in rarer cases sRNAs interact with proteins (such as RNA polymerase) to indirectly regulate the expression of target genes (e.g., 6S RNA; (Barrick, Sudarsan, Weinberg, Ruzzo & Breaker, 2005)).

The regulatory advantage of sRNAs is their ability to convey sequence-specific signals (like a zip code) to receptive targets, while requiring less genomic sequence and correspondingly lower metabolic costs than proteins (Croft, Lercher, Gagen & Mattick, 2003). The number of global protein regulation systems such as two-component regulatory systems and sigma factors are markedly reduced in open ocean microorganisms, as a result of their compact genomes presumably due in part to adaptation to their oligotrophic marine environment (Dufresne et al., 2003; Giovannoni et al., 2005b; Steglich et al., 2008). For example, only two sigma factors and four two-component regulatory systems were found in the completely sequenced genome of *Pelagibacter* strain HTCC1062 (Giovannoni et al., 2005b). The reduced number of protein regulators is correlated with the reduced biological complexity of marine microbes, but also leaves open the possibility for alternative regulatory mechanisms such as those mediated by sRNAs. In addition, sRNAs have been identified in hyper-variable genomic regions (termed genomic islands) that are postulated as hotspots for horizontally acquired genes (Padalon-Brauch et al., 2008; Sridhar & Rafi, 2007; Steglich et al., 2008). This suggests that sRNAs might be important for regulation and proper functioning of heterologous genes. Additionally, sRNA regulators are relatively convenient to co-transfer with target genes and in theory will increase the possibility of fixation of such newly acquired genes because these genes would already contain the regulatory sequences that function in the new host. Testing this hypothesis will expand our understanding of the theory that genomic islands are tightly involved in the ecology and niche adaptation of planktonic microbes just as in pathogenic microbes (Coleman et al.,

2006).

The identification and functional characterization of microbial regulatory sRNAs has been primarily restricted to a few model microorganisms and laboratory-based experimental systems (Silvaggi et al., 2006; Steglich et al., 2008; Vogel et al., 2003). As a consequence, relatively little is known about the broader diversity, expression, and regulatory targets of microbial sRNAs in the natural microbial world. The size of sRNAs (50-500 bp) makes next-generation sequencing ideal for discovery of highly expressed novel sncRNAs in nature. In Chapter 5, I describe a custom pipeline for the identification and characterization of naturally occurring sRNAs, some known and many others putative.

Model systems: bridging cultured isolates and wild populations. *Prochlorococcus* and *Pelagibacter,* the most abundant phototrophic and heterotrophic bacterium in the open ocean, respectively, are good examples of model systems for ground-truthing metatranscriptomics data, as well as integrating and leveraging findings from lab studies and meta-analysis. For example, Maureen Coleman compared microarray data for *Prochlorococcus* culture over a diel cycle (Zinser et al., 2009) to metatranscriptomic data derived from natural *Prochlorococcus* cells at different times of a day (Chapter 3), and found remarkably good correlation between *Prochlorococcus* gene expression patterns from same phase of the diel cycle, regardless of the data platform (Coleman, PhD thesis). Chapter 5 of this thesis provides another example, where I identified in a set of metatranscriptomic data a class of glycine riboswitches (a type of regulatory RNA; (Breaker, 2008)), expressed by putative *Pelagibacter*-like cells in the open ocean water column. Meanwhile, *Pelagibacter ubique* HTCC1062 culture was experimentally shown to use one of the glycine riboswitches to sense intracellular glycine level and to regulate its carbon usage for biosynthesis and energy (Tripp et al., 2008). These two examples highlight the value of well-established model systems in helping interpreting field data on their naturally occurring counterparts.

On the other hand, metatranscriptomic studies provide insights into activities of ecologically important microbes whose biology is less understood, in a general sense or under environmental conditions that have not been tested in the lab. A good illustration of the former scenario is a recently published paper by McCarren *et al* (Appendix A), where dissolved organic matter (DOM) amendment to a natural microbial community points to successional responses of

Alteromonas and *Methylophaga* populations. This has led to hypotheses of resource partitioning and synergistic interactions in degrading DOM by these organisms, which can now be tested via lab culture experiments.

Caveats and challenges

Admittedly, as with any methodology, the metatranscriptomic approach is not perfect. While it has provided an unprecedented opportunity for accessing microbial gene expression *in situ*, we need to understand its caveats and challenges in order to make sensible data interpretation and extrapolation.

Reproducibility and cross-laboratory comparison. From sample collection to final cDNA sequencing, metatranscriptomic protocols are conceptually straightforward but practically complicated and laborious. The procedure usually takes > 1 week to complete (personal experience). The length and steps of the procedure (Figure 2) inevitably raises question of how reproducible this approach is. Stewart *et al* has shown that technical reproducibility is remarkably good (Stewart et al., 2010), but less is known about reproducibility across sequencing platforms (GS 20, GS FLX, GS Titanium, Illumina, etc.), and among various laboratories. Along the same lines, studies that centrally and comprehensively compare (parts of) metatranscriptomic protocols such as the one led by He *et al* (He et al., 2010b) are in great need as they are important for setting up standards for cross laboratory comparisons.

Sequencing depth. Due to the great richness and variable evenness of microbial species found in most natural systems, as well as the high, uneven representation of transcripts from central metabolic pathways, metatranscriptomic sequencing coverage is still shallow at best. As a consequence, the majority of the transcript pool is represented by low abundance reads with little statistical confidence (e.g., singletons), albeit these may well contain important information.

Relative vs absolute. Conceptually, the least biased metatranscriptomic study would involve absolute quantification of RNA molecules in a microbial population, and directly compare results between experiments or samples. Recently, Gifford *et al* have developed an internal standard approach in an attempt to measure absolute transcript abundance in environmental samples (Gifford, Sharma, Rinta-Kanto & Moran, 2010). However, uncertainties

remain in this approach to claim "absolute" quantification, the most significant uncertainty being the unknown relative efficiency of recovery and emulsion PCR of the standard transcript. "Meta-omics" approaches have so far inevitably relied on relative quantification, which may introduce biases in comparative studies, because changes in the abundance of some transcripts would affect the relative abundance of other transcripts whose absolute abundance have not changed. However, the change needs to be dramatic in order to affect the relative abundance of other transcripts, as such effect is universal to the rest RNA pool, minimizing potential bias against any one single RNA type. Thus in many (if not most) cases, changes in relative transcript abundance will, in fact, reflect changes in the expression of specific genes.

Transcripts vs proteins. Given the complex, nonlinear relationship between gene expression, protein expression and biochemical function, the transcript profiles need to be carefully interpreted in the context of other supporting data. Reasonably good correlation between transcriptomes and proteomes, especially for transcripts and peptides in higher abundance, has been observed in several model organisms (Corbin et al., 2003; Eymann, Homuth, Scharf & Hecker, 2002; Scherl et al., 2005). Nonetheless, transcript abundance will not always correlate directly with cognate protein levels, and the kinetics that relate expression to phenotype varies among different transcript classes (Steunou et al., 2008; Jacob Waldbauer, PhD thesis). Transcript profiling should hence be viewed as a global but preliminary indicator of changing biology and environmental conditions, that cannot fully substitute for detailed functional and ecological analyses of candidate microorganisms or genes.

Bioinformatic challenges. As higher sequencing coverage is becoming a sought-after feature, metatranscriptomic-centered studies face several informatics challenges, from the development of efficient methods to store, retrieve and analyze large amounts of data, to the efficient communication and presentation of findings from such large data sets. Particularly, the quality of metatranscriptomic researches relies heavily on the bioinformatic infrastructure available, including the capacity to generate high quality gene annotations, statistical inferences, and metadata integration.

The structure of this thesis

Chapter 2 describes the methodological development of the first marine microbial metatranscriptomic study that used next-gen sequencing. I cross-validated the method using microarray data of the *Prochlorococcus* cultures. Furthermore, I carried out a pilot study applying this approach in studying the community transcriptome of a bacterioplankton sample in the open ocean photic zone.

In **Chapter 3**, I extended the metatranscriptomic survey to four bacterioplankton samples along the vertical water column in the open ocean, and integrated those with metagenomic survey of the same set of samples. I performed comparative analyses to describe genomic content and transcriptomic composition of microbial assemblages in these distinct environmental settings.

In addition to surveying *in situ* microbial gene expression, deep sequencing-based metatranscriptomics provides a useful approach for monitoring instantaneous responses of microbes under controlled perturbation experiments. In **Chapter 4**, I simulated a deep water mixing event in a microcosm setting, and applied metatranscriptomics over the course of 27 hours to monitor community structural and transcriptional dynamics.

Chapter 5 describes the identification and characterization of highly expressed known and novel small RNAs (sRNAs) in metatranscriptomic data sets. In particular, I introduced for the first time a bioinformatic pipeline tailored for sRNA studies using metatranscriptomic data. This study and those alike provide important insights into the dynamic sRNA species and their specific interplay with community taxonomic structure, microbial activity and environmental conditions, laying foundation for future biochemical and genetic characterization of identified sRNAs.

Finally, in **Chapter 6** I conclude and integrate our findings from the 4 interrelated studies, and point out future research directions. I integrated metatranscriptomic and metagenomic approaches, in natural settings as well as in controlled perturbation experiments, to address questions at various levels such as the following. Which taxa of marine Bacteria and Archaea are most dominant or functionally important in particular ocean provinces or depth strata? What are the common versus habitat-specific microbial metabolic pathways, and how do

they vary with different communities and environments? Can we detect expression signals of low-abundance populations that may nevertheless play important ecological roles? Can we detect molecular-level regulatory interactions in the community transcriptomes? As a whole, this thesis provides a new set of insights towards understanding the expression and regulation of microbial functions, as well as the environmental factors (biotic and abiotic) that influence microbial assemblage dynamics in the open ocean.

Figures

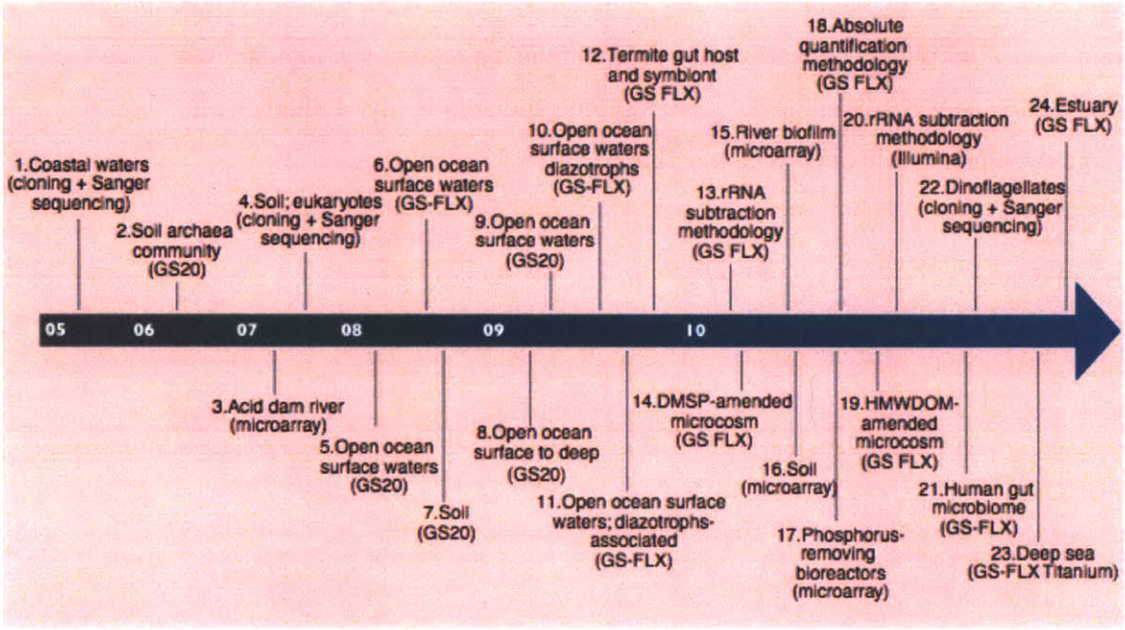


Figure 1. Timeline of publications on microbial community transcriptomics (metatranscriptomics). Targeted samples in these studies were bacterial and archaeal communities unless otherwise specified. Sequencing platforms were specified in the parentheses. DMSP: dimethylsulfoniopropionate. HMWDOM: high molecular weight dissolved organic matter. The references are listed below. 1. (Poretsky et al., 2005); 2. (Leininger et al., 2006); 3. (Parro et al., 2007); 4. (Bailly et al., 2007); 5. (Frias-Lopez et al., 2008); 6. (Gilbert et al., 2008); 7. (Urich et al., 2008); 8. (Shi et al., 2009); 9. (Poretsky et al., 2009); 10. **(Hewson et al., 2009a)**; 11. **(Hewson et al., 2009b)**; 12. (Tartar et al., 2009); 13. (Stewart et al., 2010); 14. (Vila-Costa et al., 2010); 15. **(Yergeau, Lawrence, Waiser, Korber & Greer, 2010)**; 16. (McGrath et al., 2010); 17. **(He et al., 2010a)**; 18. (Gifford et al., 2010); 19. (McCarren et al., 2010); 20. **(He et al., 2010b)**; 21. (Turnbaugh et al., 2010); 22. **(Lin, Zhang, Zhuang, Tran & Gill, 2010)**; 23. (Wu et al., 2010); 24. **(Hollibaugh, Gifford, Sharma, Bano & Moran, 2010)**.

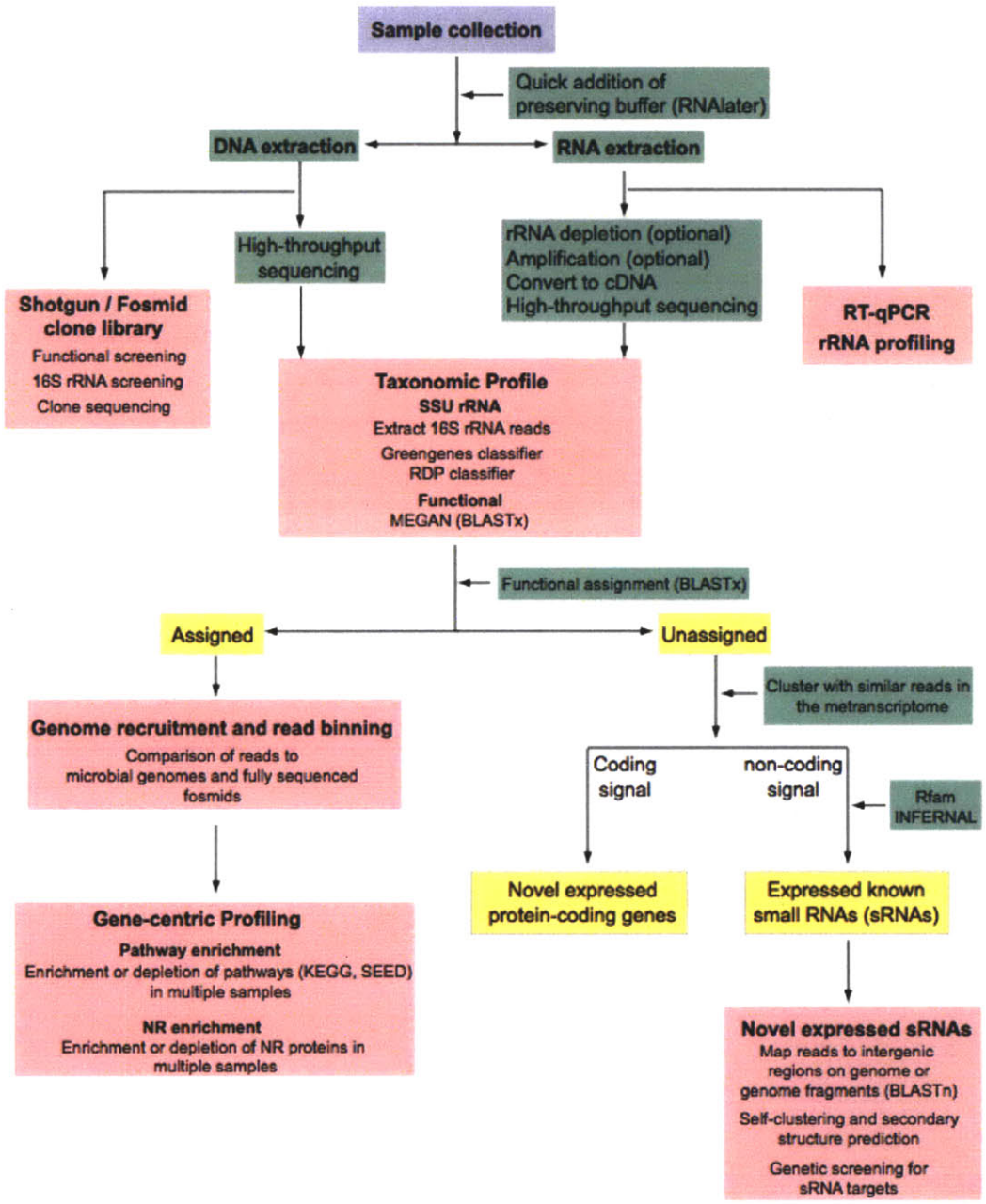


Figure 2. A non-exhaustive pipeline for next-gen sequencing-based metatranscriptomic studies. This pipeline is based on ongoing metatranscriptomic research in the DeLong lab, and thus does not necessarily represent experimental and analytical procedures undertaken by other researchers.

CHAPTER TWO

Microbial community gene expression in ocean surface waters: methodology and a pilot study of microbial metatranscriptomics

Jorge Frias-Lopez*, Yanmei Shi*, Gene W. Tyson, Maureen L. Coleman, Stephan C. Schuster, Sallie W. Chisholm, and Edward F. DeLong

*These authors contributed equally to this work.

This chapter is presented, with slight formatting modification, as it appeared in *Proc. Natl. Acad. Sci. USA* **105**, 3805-3810 (2008). Corresponding supplementary information is appended.

Chapter 2: Microbial community gene expression in ocean surface waters: methodology and a pilot study of microbial metatranscriptomics

Abstract

Metagenomics is expanding our knowledge of the gene content, functional significance, and genetic variability in natural microbial communities. Still, there exists limited information concerning the regulation and dynamics of genes in the environment. We report here global analysis of expressed genes in a naturally occurring microbial community. We first adapted RNA amplification technologies to produce large amounts of cDNA from small quantities of total microbial community RNA. The fidelity of the RNA amplification procedure was validated with Prochlorococcus cultures, and then applied to a microbial assemblage collected in the oligotrophic Pacific Ocean. Microbial community cDNAs were analyzed by pyrosequencing, and compared to microbial community genomic DNA sequences determined from the same sample. Pyrosequencing-based estimates of microbial community gene expression compared favorably to independent assessments of individual gene expression using quantitative PCR. Genes associated with key metabolic pathways in open ocean microbial species, including genes involved in photosynthesis, carbon fixation, and nitrogen acquisition, and a number of genes encoding hypothetical proteins, were highly represented in the cDNA pool. Genes present in the variable regions of Prochlorococcus genomes were among the most highly expressed, suggesting these encode proteins central to cellular processes in specific genotypes. Although many transcripts detected were highly similar to genes previously detected in ocean metagenomic surveys, a significant fraction (~ 50%) were unique. Thus, microbial community transcriptomic analyses revealed not only indigenous gene- and taxon-specific expression patterns, but also new gene categories, undetected in previous DNA-based metagenomic surveys.

Introduction

Cultivation-independent genomic approaches have greatly advanced our understanding of the ecology and diversity of microbial communities in the oceans (DeLong & Karl, 2005; Giovannoni & Stingl, 2005). Metagenomic methods have been applied in a variety of microbial habitats, and have led to the discovery and characterization of new genes and gene products from uncultivated microorganisms (Béjà, Spudich, Spudich, Leclerc & DeLong, 2001), assembly of whole genomes from community DNA sequence data (Tyson et al., 2004), and comparisons of community gene content among diverse microbial assemblages (Angly et al., 2006; Coleman et al., 2006; DeLong et al., 2006; Gill et al., 2006; Tringe & Rubin, 2005; Turnbaugh et al., 2006; Tyson et al., 2004; Venter et al., 2004). Recently, a very large metagenomic sampling survey was conducted in ocean surface waters, doubling the number of predicted protein sequences in

public databases (Rusch et al., 2007). All currently available data suggest that gene and protein "sequence space" still remain largely under sampled.

At the same time, studies of cultured members of the microbial community, such as *Prochlorococcus*, are helping to further link the ecology of genes and the ecology of organisms (Coleman & Chisholm, 2007). From the considerable *Prochlorococcus* diversity observed in metagenomic datasets clear structure has emerged, including clusters of sequence similarity and chromosomal hotspots for rearrangements (Coleman et al., 2006; Rusch et al., 2007; Venter et al., 2004). Meanwhile, laboratory studies have described physiological differentiation among isolates (Moore & Chisholm, 1999; Moore, Ostrowski, Scanlan, Feren & Sweetsir, 2005), and field surveys have documented the distribution of ecotypes in the oceans (Johnson et al., 2006). These cross-scale comparisons provide a useful approach in which taxon specific metagenomic information can be embedded and understood in the context of ecological and physiological data.

Given current research trends, it seems likely that metagenomic datasets will continue to grow rapidly, and will soon dwarf whole genome sequence datasets derived from cultivated microorganisms. The nature, size and complexity of this information present formidable challenges to analyses and interpretation. In addition, while these data provide information about genome content, there is no clear indication of gene expression or expression dynamics. Whereas techniques like quantitative PCR can be used to quantify gene expression in natural samples, these are limited usually to measurement of a small number of known genes. What fraction of the many new genes discovered in metagenomic datasets are actually expressed? Of the many hypothetical genes present, which are significantly expressed, and what is their function? What are the dynamics and time scales for gene expression in different microbial species, gene suites, and environments?

Measuring bacterial and archaeal gene expression in the wild has been challenging. The half-life of mRNA is short (Andersson et al., 2006; Selinger, Saxena, Cheung, Church & Rosenow, 2003) and therefore microbial biomass must be harvested rapidly. Furthermore, mRNA in bacteria and archaea usually comprises only a small fraction of the total RNA. A number of methods to overcome these challenges have recently been developed. In one approach, rRNA subtraction was used in combination with randomly primed reverse transcription PCR, to generate microbial community cDNA for cloning and downstream

sequence analysis (Poretsky et al., 2005). While preliminary results were encouraging, relatively large sample volumes (~ 10 liters) and long sample collecting times were required. Linear RNA amplification methods have been widely used to study gene expression in eukaryotic tissues (Dafforn et al., 2004; Feldman et al., 2002; Moll, Duschl & Richter, 2004; Schneider et al., 2004), but this generally requires the presence of a polyadenylated tail on the 3' end of the mRNA, which is not characteristic of bacterial nor archaeal mRNA. To overcome this problem, Wendisch *et al* (Wendisch et al., 2001) developed a method for the polyadenylation of bacterial messenger RNA using *E. coli* poly (A) polymerase, which allowed preferential isolation of bacterial mRNA from rRNA in crude extracts. This approach has been adapted in a commercially available kit (MessageAmp II-Bacteria Kit, Ambion, Austin, TX), which couples microbial RNA polyadenylation with a linear amplification step using T7 RNA polymerase (Vangelder et al., 1990). Polyadenylation-dependent RNA amplification approaches have been used in studies of cultured microbes using single genome microarrays (Moreno-Paz & Parro, 2006; Rachman, Lee, Angermann, Kowall & Kaufmann, 2006). We adapted this approach to enable the synthesis of microbial community cDNA, from small amounts of mixed population microbial RNA. Specifically, following *in vitro* enzymatic polyadenylation of nanogram quantities of RNA (Wendisch et al., 2001), the RNA was linearly amplified using T7 RNA polymerase (Vangelder et al., 1990), and the amplified RNA converted to cDNA. The cDNA was then directly sequenced by pyrosequencing, avoiding the need to prepare clone libraries, and their associated biases (Huse, Huber, Morrison, Sogin & Welch, 2007; Margulies et al., 2005). By sequencing both genomic DNA and cDNA from the same sample it was possible to normalize the abundance of cDNA copies relative to corresponding gene copy numbers in the community DNA pool.

We report here the application, validation, and field-testing in the North Pacific Subtropical Gyre (Karl & Lukas, 1996), of these methodologies for studying microbial community gene expression. We used the technique to analyze the expression of genes across the entire microbial community, to assess the taxonomic origins of the expressed genes, and to examine gene expression in *Prochlorococcus*, the dominant phototroph in the surface waters at this site. Genes from *Prochlorococcus* are highly represented in metagenomic databases (DeLong et al., 2006; Rusch et al., 2007; Venter et al., 2004), and extensive genomic and transcriptomic data exists from culture studies (Coleman et al., 2006; Dufresne et al., 2003;

Holtzendorff et al., 2001; Martiny, Coleman & Chisholm, 2006; Rocap et al., 2003; Tolonen et al., 2006), and so were useful in guiding the interpretation of field observations.

Materials and methods

Sampling

Seawater was collected at the Hawaii Ocean Time Series (HOT) station ALOHA (22°44′N, 158°2′W), 75 m depth, on March 9, 2006, 03:30 a.m. local time. Hydrocasts for sampling and hydrographic profiling were conducted using a conductivity-temperature-depth (CTD) rosette water sampler equipped with 24 Scripps 12-l sampling bottles aboard the R/V Kilo Moana. Continuous vertical profiles of physical and chemical parameters were thus recorded. DNA and RNA extraction, processing and sequencing are detailed in the Supplementary Information.

RNA amplification and cDNA synthesis

~5 μl RNA (~ 100 ng total) was amplified using MessageAmp II-Bacteria Kit (Ambion, Austin, TX) following manufacturer's instructions. Briefly, the method is based on polyadenylation of the 3'-end of total RNA. The A-tailed RNA is reverse transcribed primed with an oligo(dT) primer containing a T7 promoter sequence and a restriction enzyme (BpmI) recognition site sequence (T7-BpmI-$(dT)_{16}$VN), then double-stranded cDNA is synthesized. Finally, the cDNA templates are transcribed in vitro (37 °C for 6 hours), yielding large amounts of antisense RNA (aRNA; ~ 1000 fold amplification). The aRNA is polyadenylated and further reverse transcribed to cDNA using SuperScript™ Double-Stranded cDNA Synthesis Kit (Invitrogen, Carlsbad, CA). Finally, ~ 2 μg of cDNA is digested with BpmI, purified, and used for pyrosequencing.

Pyrosequencing

DNA and cDNA libraries were constructed as previously described (Margulies et al., 2005; Poinar et al., 2006) and sequenced using a Roche GS20 DNA sequencer. A full run of the sequencer yielded 45,380,301 bps from 414,323 reads (110 bp average length) from the DNA library, and 14,675,424 bps from 128,324 reads (114 bp average length) from the cDNA library

(Table 1). The lower number of cDNA library reads may be due to shorter cDNA fragments and highly polymeric sequences resulting from inefficient removal of poly(A) tails introduced during mRNA amplification. To pass GS20 quality filters, flowgrams for each read require at least 84 flows (21 cycles, or approximately 50 bps) and < 5% of flows with ambiguous bases (N) and < 3% of flowgram values between 0.5-0.7 (GS20 Data Processing Software Manual).

Analysis of metagenomic GS20 DNA and cDNA data

DNA and trimmed non-RNA cDNA reads were compared to the NCBI non-redundant protein (NCBI-nr; as of March 28, 2007) and Global Ocean Survey (GOS) peptides databases using BLASTX (Altschul, Gish, Miller, Myers & Lipman, 1990). Top BLASTX hits with bit score > 40 were used to assign DNA and cDNA reads to GOS peptides and NCBI-nr proteins (Table 1). Reads assigned to GOS peptides were linked to GOS protein clusters and associated GO, Pfam, and TIGRfam annotations (if available). Additional details are provided in Supplementary Information.

Results and Discussion

Assessing the fidelity of bacterial mRNA amplification

We tested the fidelity of the RNA amplification technique using *Prochlorococcus* cultures and custom designed Affymetrix arrays (see Supplementary Information) (Martiny et al., 2006). Levels of gene expression measured from the amplified *Prochlorococcus* RNA compared favorably with those of unamplified RNA for protein coding genes (r^2 between 0.85 and 0.92; Figure S1), and the results were highly reproducible (r^2 between 0.94 and 0.99 for biological replicates; Figure S2). Linearly amplified RNA also revealed the same physiologically relevant changes in gene expression, as did unamplified RNA in an experiment designed to examine the response of strain MIT9313 to phosphate starvation (Figure S3) (Martiny et al., 2006). Both amplified and unamplified RNA identified the same four genes, all involved in phosphate acquisition, as highly up-regulated under P-starvation. In contrast to this high fidelity for mRNA, ribosomal RNA (rRNA) transcripts were consistently underrepresented in amplified versus unamplified RNAs (Figure S4), reflecting a preferential polyadenylation of mRNA, consistent with previous reports of this polyadenylation bias in crude extracts (Wendisch et al.,

2001), and with the known inefficiency of amplification of molecules with a high degree of secondary structure (von Wintzingerode, Göbel & Stackebrandt, 1997).

Field-testing microbial gene expression profiling in the open ocean

As a field test, we analyzed a picoplanktonic sample collected from 75 m depth at the well-characterized Hawaii Ocean Time-series station ALOHA, in the North Pacific Subtropical Gyre (Karl & Lukas, 1996). Since metagenomic analyses have already been performed at this site (Coleman et al., 2006), and the cyanobacterium *Prochlorococcus* comprises a large fraction of its microbial communities (Campbell, Liu, Nolla & Vaulot, 1997; Campbell, Nolla & Vaulot, 1994), databases exist to facilitate the interpretation of our field results. Since the detection frequency of any given transcript in the community depends on the abundance of transcript-bearing cells (reflected by gene abundance in community genomic DNA), and the average number of transcripts per cell (reflected in their cDNA abundance), we recovered sequence data from both cDNA and genomic DNA in the same sample. This allows the representation of specific cDNA classes relative to their occurrence in the genomic DNA pool, i.e. an estimate of relative expression per gene copy.

The diversity of sequences captured in the cDNA and DNA reads (Table 1) was determined by comparing all sequences to the NCBI-nr protein database, and to predicted peptides from the recent Global Ocean Sampling (GOS) metagenomic dataset (Yooseph et al., 2007). The number of cDNA and DNA reads with significant database matches (bits score > 40; Figure S4) was higher with GOS peptides, than with the NCBI-nr database. This was expected, because the GOS data are derived from similar microbial communities and contain a larger number of total protein sequences. The enrichment in GOS matches over NCBI-nr matches was much greater for the cDNA library (~3 fold) compared to the DNA library (~1.4 fold) (Table 1). The fraction of reads matched in the cDNA however, was still relatively low (43% of total reads) compared to the DNA library (70% of reads). The large proportion of unmatched cDNA reads may in part reflect the presence of novel, rare genes, not detected in the GOS metagenomic survey, that nevertheless contribute significantly to the microbial community expression profile.

To corroborate the results we selected a suite of genes and performed quantitative reverse transcription-PCR (RT-qPCR) and qPCR on the same RNA and DNA samples analyzed by pyrosequencing (Supplementary Methods, Table S1, and Figure S6). Three different gene

expression classes were investigated: 1) genes shared in both genomic DNA and cDNA sequence datasets, but with higher relative frequency in the cDNA pool, 2) genes present in both genomic DNA and cDNA datasets but with lower relative frequency in the cDNA pool, and 3) genes detected in the cDNA but not in the genomic DNA sequence dataset. The calculated RT-qPCR/qPCR ratios followed the same trends as gene expression patterns inferred from cDNA/DNA pyrosequencing analyses (Figure S6). In some cases, the RT-qPCR/qPCR analysis appeared more sensitive for detecting a broader range of gene expression patterns. For example, genes found only in the cDNA sequence dataset were detected by qPCR in both RNA and DNA samples. This likely reflects the limited extent of sampling depth of the DNA pyrosequencing relative to indigenous genetic complexity.

To evaluate the protein family representation in our dataset and to functionally categorize genes, reads from both cDNA and DNA libraries were assigned to GOS protein clusters using BLASTX. DNA reads were assigned to 35,178 different GOS protein clusters, while cDNA reads were assigned to 4,376 clusters. There were 2,654 clusters that had both DNA and cDNA reads (Figure 1). The smaller number of cDNA assignments is in part because the total number of cDNA reads was only one-eighth the number of DNA reads, after removing rRNA sequences. Another factor likely responsible for the decreased number of high quality sequence reads in the cDNA relative to genomic DNA, includes the inefficient enzymatic removal of the poly (A) tail produced during the amplification of the mRNA. These homopolymers cause a significant number of sequences to be filtered out during processing due to lower quality scores, low flow counts, and carry forward (premature incorporation of bases due to incomplete flushing) (see Materials and Methods; (Huse et al., 2007)). Nevertheless, 40% of the cDNAs contained in GOS clusters (referred to as cDNA-unique clusters hereafter) did not overlap with those in the DNA library, suggesting that the full diversity of sequences was under-sampled in both the DNA and cDNA pools. This is supported by rarefaction analysis, showing a near linear increase in the rate of recovery of GOS protein clusters with increasing number of sequence reads for both cDNA and DNA (Figure S7). This finding is consistent with other large-scale metagenomic surveys that showed no sign of sequencing saturation for similar marine microbial communities (Sogin et al., 2006; Yooseph et al., 2007).

To maximize functional genomic information drawn from the data, the 2,654 GOS

protein clusters (protein families) that were represented in both the DNA and cDNA libraries were analyzed further, calculating the number of cDNA reads matching a given GOS protein cluster, divided by the number of corresponding DNA reads in the same cluster (see Material and Methods) — the 'cluster-based expression ratio'. This approach allowed us to bypass the difficulties associated with traditional annotation of short pyrosequencing reads (average trimmed length of ~96 bp), which would have segmented the reads into many apparently unrelated, non-overlapping clusters, even though they were potentially derived from the same gene. This level of analysis allows us to look at the expression profile of the microbial community at the level of protein family, without losing the resolution inherent in the data.

The 2,654 shared GOS protein clusters were categorized based on their abundance in the DNA library (low, medium, high and extremely high; Figure S8). Protein clusters with the highest cluster-based expression ratios (up to 10^3 higher than the average ratio) tended to fall into the low DNA abundance category (Figure 1B). This observation, together with apparent high expression levels in cDNA-unique clusters, suggested the presence of actively transcribed genes that are relatively low in abundance in the total community. Interestingly, these highly expressed protein clusters consist mostly of hypothetical proteins that are found only in the GOS peptide database (Figure 1; Table S2). The high degree of sequence similarity (up to 100%; average 89.5%) between these GOS-only hypothetical protein matches and the cDNA reads validates the GOS gene predictions and confirms that these genes are actively expressed *in situ*. Conversely, the DNA-unique clusters are composed of protein families that are well represented in current protein databases (e.g., NCBI-nr and fully sequenced microbial genomes; Figure 1; Table S3). This contrast further illustrates that cDNA analysis can capture novel genes, with potentially important functions, that have escaped detection even in the largest metagenomic DNA survey conducted to date.

Highly expressed gene categories in known metabolic pathways

Expression patterns of environmentally diagnostic genes can provide significant insight into microbial processes active in the environment. For example, genes involved in microbial phototrophy — e.g. oxygenic and anoxygenic photosynthesis and photoheterotrophy — were among the most highly expressed classes in cluster-based expression ratios (Figure 1B and see *Prochlorococcus* section below) even though the sample was collected three hours before

sunrise.

In the case of genes related to oxygenic photosynthesis, Ribulose bisphosphate carboxylase (RuBisCo) large subunit (*rbc*L) homologs, encoding subunits of the key enzyme in the Calvin Cycle carbon fixation enzyme were among the highly expressed genes in the sample (Figure 1B). Expression levels of this gene were on a par with those of glutamine synthase (GS), suggesting high expression levels of this key enzyme in nitrogen metabolism that is found in all microorganisms. RuBisCo and GS gene copies were present in comparable numbers in the microbial genomic DNA of our sample, in contrast to the recently reported GOS datasets, where relatively low numbers of the *rbc*L gene were identified, relative to GS (Yooseph et al., 2007). With respect to alternative forms of phototrophy, several protein clusters associated with aerobic, anoxygenic phototrophy showed extremely high cluster-based expression ratios (Figure 1B). These proteins include light-harvesting protein beta chain (PufB), photosynthetic reaction center cytochrome C subunit (PufC), and chlorophyllide reductase subunit Y (BchY), that all appear to be derived from Alphaproteobacteria closely related to *Roseobacter* species (Oz, Sabehi, Koblízek, Massana & Béjà, 2005). Although these correspond to relatively low abundances in the DNA libraries, their high expression levels support the potential ecological importance of aerobic anoxygenic phototrophy to microbial species in the open ocean.

Another important family of proteins involved in phototrophy are the proteorhodopsins, a group of membrane proteins that function as a light-driven proton pump (Béjà et al., 2001). Proteorhodopsin (PR) genes were not only abundant in community genomic DNA, but also were among the most highly expressed genes in the cDNA pool (Figure 1). Preliminary taxonomic assignments suggest that the expressed PR genes were derived from diverse microbial taxa, supporting their general ecological significance in planktonic microbial communities (Béjà et al., 2001; Sabehi et al., 2005). Heterologous expression experiments have confirmed the ability of PR to function as a proton pump, and enable photophosphorylation in *E. coli* (Béjà et al., 2001; Martinez, Bradley, Waldbauer, Summons & DeLong, 2007). Moreover, some PR-containing bacteria display enhanced growth rates and cell yields in the presence of light (Giovannoni et al., 2005a; Gómez-Consarnau et al., 2007).

Putative taxonomic origins of expressed genes

Metatranscriptomic analyses can, in principle, be used to associate specific microbial taxa

with *in situ* expression dynamics. However, phylogenetic inference based on protein-coding genes is highly dependent on a given gene's conservation across taxa, the depth of taxonomic sampling, taxon richness and evenness in the sample, and sequence read length. Further, taxonomic inferences also have the potential to be confused by horizontal gene transfer events (Boucher et al., 2003). With these caveats in mind, we performed a preliminary taxonomic assessment of DNA and cDNA reads using MEGAN (Huson, Auch, Qi & Schuster, 2007), software that assigns putative taxonomic origins based on BLAST outputs, and NCBI taxonomic hierarchy. Not surprisingly, based on their known abundance in the wild and their abundance in the genomic databases, the genus *Prochlorococcus*, and Alphaproteobacteria (genus *Pelagibacter*) were the two most highly represented taxonomic groups in both DNA and cDNA libraries (Figure 2 and Table S4). Another noteworthy observation was the detection of expressed genes of viral origin, suggesting there was active viral infection occurring in cells *in situ* in the sample we analyzed (Figure 2 and Table S4). The most common viral transcripts were related to the major capsid protein of myoviridae. Previous metagenomic analyses reported a high viral abundance in the cellular fraction from the same depth and site (DeLong et al., 2006). For the most abundant groups, there was general agreement between the taxonomic origins of sequence reads in the DNA and cDNA datasets.

Evaluating gene expression in a naturally occurring *Prochlorococcus* assemblage

As the most abundant oxygenic phototroph in these waters (Campbell et al., 1994), and with 12 complete genome sequences available, *Prochlorococcus* provides a unique opportunity for in-depth analysis of gene expression of a single microbial group *in situ*. Because of the extensive genomic database for this genus, sequence reads can be assigned specifically to well-annotated genome sequences, and in some cases to the specific ecotypes expressing these genes.

The vast majority (over 90%) of putative *Prochlorococcus* reads shared highest sequence similarity with strains MIT9301, AS9601, and MIT9312, all representatives of the high light-adapted eMIT9312 ecotype (Rocap, Distel, Waterbury & Chisholm, 2002). This result (data not shown) is consistent with depth-specific ecotype abundance data based on quantitative PCR analysis of the rRNA internally transcribed spacer (ITS) region (Johnson et al., 2006). Our current analysis using short pyrosequencing sequence reads from both DNA and cDNA therefore support ecotype distributions inferred from independent analyses using a single taxonomic

marker, the ITS.

Observed frequencies of the putative *Prochlorococcus* cDNA sequences reflect which genes are the most highly expressed in the *Prochlorococcus* assemblage sampled. These highly expressed genes include ammonium uptake (*amt*), photosynthesis (*psa*AB), and carbon fixation (*rbc*L) genes, pointing to key biogeochemical processes being driven, in part, by *Prochlorococcus* (Figure 3A; Table S5). Two of the top twenty most highly expressed *Prochlorococcus* genes were hypothetical proteins: P9301_11381, which has orthologs only in the other MIT9312-like genomes (AS9601, MIT9312, and MIT9215), and P9301_07111, which has no orthologs in other *Prochlorococcus* genomes (but is paralagous to P9301_04361) (Table S5). High-level expression of hypothetical proteins has previously been observed in *Prochlorococcus* under nutrient limitation in laboratory experiments (Martiny et al., 2006; Tolonen et al., 2006). The current data indicate the potential relevance of these proteins to *Prochlorococcus* in its native environment. When a gene-length correction is applied (see Materials and Methods; Figure S9; Table S5), additional hypothetical proteins (P9301_03541, P9301_02451) with high per-copy transcript abundance appear to be rare in the population, but are highly expressed.

The *Prochlorococcus* core genome (i.e., those genes shared by all sequenced *Prochlorococcus* isolates) consists of approximately 1250 genes (Kettler et al., 2007). The "flexible" genome represents the remaining genes found in one or more genomes, and many of these variable genes are concentrated in genomic islands (Coleman et al., 2006). Using strain MIT9301 as a reference, we calculated the abundance of genes belonging to the core and flexible genomes in both the DNA and cDNA libraries. In the DNA library, all *Prochlorococcus* core genes were represented with roughly equal abundance, supporting the idea that these genes are conserved and present in single-copy in virtually every *Prochlorococcus* cell (Figure 3B). In contrast, genes belonging to the MIT9301 flexible genome had highly variable occurrence in the DNA library, suggesting that the natural population likely harbors a different suite of such genes. In the cDNA library, core genes involved in photosynthesis and carbon fixation, for instance, were highly represented, but, surprisingly, a number of genes belonging to the flexible genome, some of which are located in genomic islands in MIT9301, were also highly represented (Figure 3A, 3C). Thus some of these island genes appear to be highly expressed, corroborating

laboratory evidence, and suggesting that they are likely functionally important to naturally occurring *Prochlorococcus*. Furthermore, the majority of 'flexible' genes, as well as hypotheticals, were found in the cDNA pool and expressed at levels comparable to most other core genes, further indicating their significance in the biology and ecology of *Prochlorococcus*.

Microbial community transcriptomics: prospects and challenges

Many new challenges are associated with the interpretation of microbial gene expression patterns at the community level. These arise in part from the remarkable diversity and complexity of microbial communities in the ocean environment, the significant challenges associated with field sampling, the shortage of cultured model organisms, and the lack of comprehensive representation in metagenomic databases. Rapid collection and processing of samples for gene expression studies, for example, still presents significant challenges. While our approach employed relatively small volumes (1 liter) and short filtration times (< 15 min.), there still remains significant room for improvement. Other factors that will influence community transcriptomic analyses include the specifics of mRNA synthesis and degradation rates, environmental conditions at the time of sampling (time of day, for example), sequence read size and target gene size, and the specific method used for gene identification and annotation. Some of these variables can be controlled or improved, and others are inherent to the specific environment or community being sampled.

It is well accepted that longer sequence reads are generally more informative, allowing more robust annotation. Side-by-side comparisons of Sanger dideoxy sequences versus pyrosequences derived from the same metagenomic samples however have been generally consistent and comparable with one another (Gill et al., 2006; Turnbaugh et al., 2006). The sequence reads in our dataset have an average size of ~ 96 bp, sufficient for general functional annotation, and in the case of *Prochlorococcus*, for assignment of reads to specific genes and ecotypes. For as yet uncultivated microorganisms, or those with fewer reference genomes available however, 100 bp may not be sensitive enough for specific gene assignment. Improvements in pyrosequencing however now produce >230-bp length reads, and in the near future will likely yield even longer, high quality sequence reads. These advances are expected to improve even more, further enabling application of microbial community transcriptomics in future studies.

Despite the caveats and potential improvements to the approach reported here, we have shown metatranscriptomic sequencing and characterization (based on amplified RNA and pyrosequencing) is sufficient to identify many expressed biological signatures (including microbial taxa, and specific protein families) in complex biological samples such as seawater. Whole community analysis relying on gene family clustering for analyses of pyrosequencing reads revealed clear patterns in community gene expression for both individual taxa, specific genes, and within protein families. Taxon-specific analyses focusing on *Prochlorococcus* provided deep insight into the most highly expressed genes among these populations. Interestingly, both in the case of the whole community as well as in the case of *Prochlorococcus*, hypothetical genes were among the most highly expressed, underlining the potential importance of these unidentified proteins. The fact that a large fraction of cDNA reads were not present in the available databases, including the GOS database, indicates that we have just scratched the surface of the microbial metabolic diversity present in the ocean.

Metatranscriptomics ((Poretsky et al., 2005), this report) and proteomics (Lo et al., 2007; Ram et al., 2005) represent two new approaches in microbial ecology that have potential to significantly leverage, apply, and extend existing microbial metagenomic datasets. The two approaches each measure a different component and dynamic of the macromolecular pool, reflecting the different regulatory controls, expression rates, and turnover kinetics of mRNAs and proteins. While transcriptomics has potential to reveal the near instantaneous responses to environmental fluctuation, proteomics more directly reflects the immediate catalytic potential of the microbial community. In conjunction with metagenomic data, these approaches offer significant promise to advance measurement and prediction of *in situ* microbial responses and activities in complex, naturally occurring or engineered microbial communities.

Table and Figures

Table 1. Characterization of the pyrosequence DNA and cDNA libraries from the microbial community analyzed in the study.

	DNAlibrary	cDNAlibraray
Total number of reads	414,323	128,324
Average length (bp)	110	114
Number of rRNA reads	5,877	67,859
Total base pairs (Mb)	45.4	14.7
Number of NCBI-nr hits[1]	205,747 (50% of reads)	7,275 (13% of reads)
Number of GOS peptide hits[1]	290,741 (70% of reads)	23,203 (43% of reads)

[1]**Only sequences whose bits score >= 40 were considered hits.**

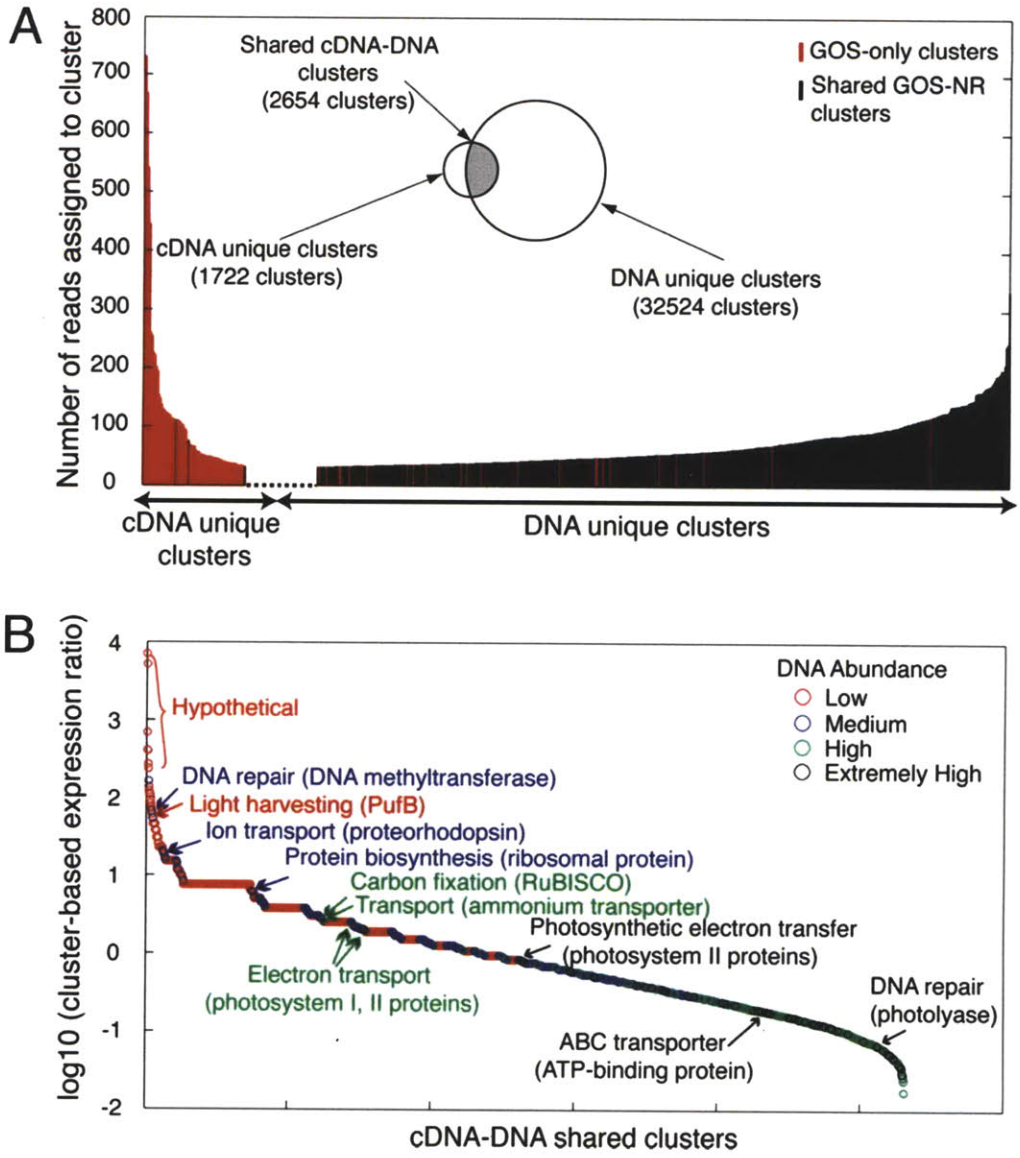


Figure 1. Community-level gene expression profile based on GOS peptide database. (*A*) GOS protein clusters with DNA or cDNA matches at bit scores ≥40 are shown in the Venn diagram. Numbers of reads assigned to GOS protein clusters, when >70, are plotted for both cDNA-unique protein clusters and DNA-unique protein clusters. GOS protein clusters shared by DNA and cDNA libraries (shaded in gray) were further illustrated in *B*. (*B*) GOS protein clusters shared by cDNA and DNA libraries were ranked by their cluster-based expression ratio (representation of each cluster in the cDNA library normalized by its representation in the DNA library). Furthermore, each protein cluster was categorized (and color-coded) according to its abundance in the DNA library. Representative protein clusters were highlighted from each category and discussed in the text.

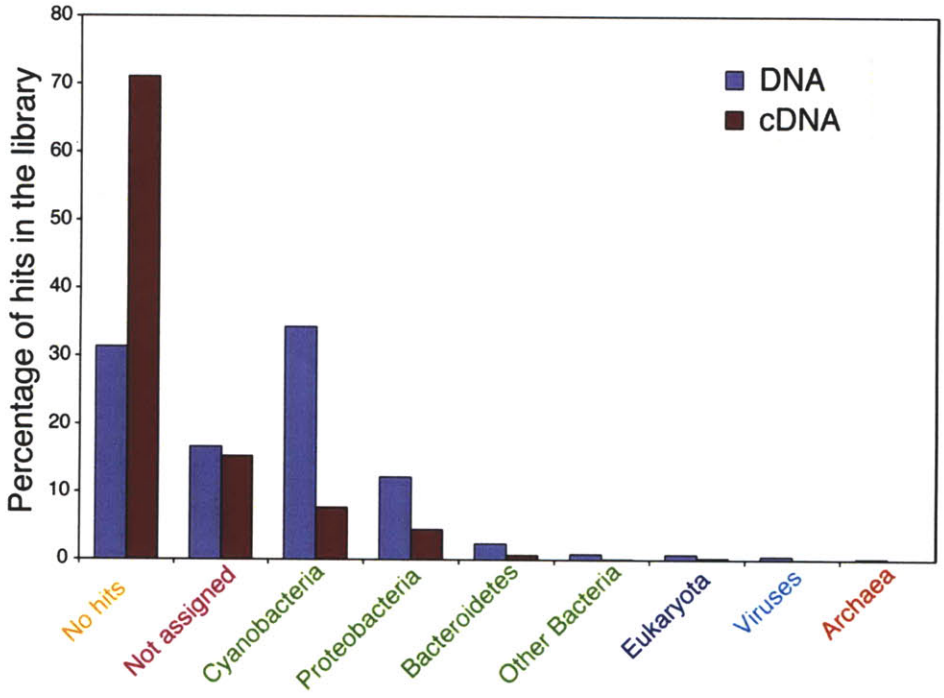


Figure 2. Distribution of different phylogenetic groups in DNA and cDNA libraries. Percentages of the different phylogenetic groups were calculated from the MEGAN analysis results at the phylum level cutoff (Table S4 shows a detailed list of the distribution of number of hits and percentages for all phyla). Not assigned reads are sequences with an NR hit but a bit score <40.

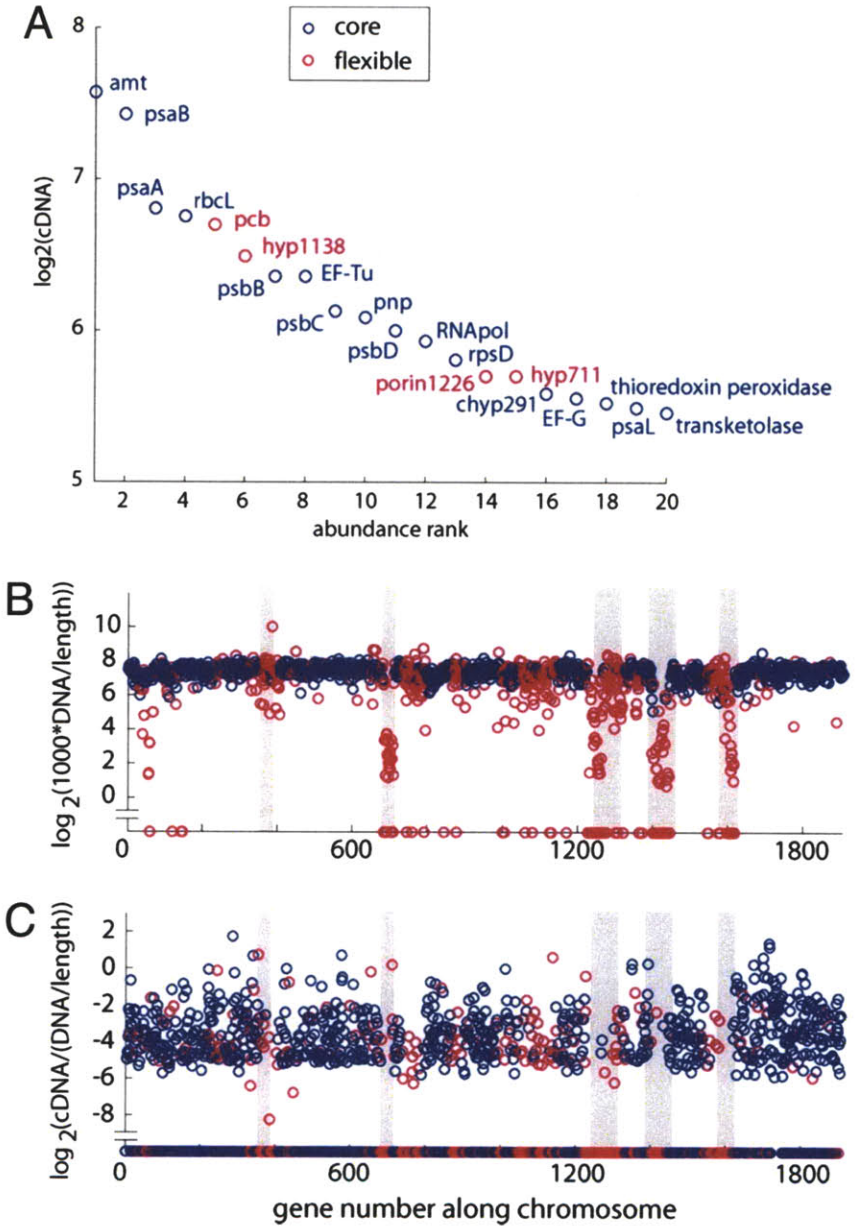


Figure 3. *Prochlorococcus* gene and transcript abundance using strain MIT9301 as a reference genome. (*A*) Rank abundance of the 20 genes with highest frequency in the raw cDNA, reflecting transcription of the entire *Prochlorococcus* population. (*B*) Frequency of DNA hits from the natural sample along the genome of MIT9301 normalized to gene length. (*C*) Frequency of cDNA hits from the natural sample normalized to the DNA values in *B*. Gray bars indicate the location of genomic islands identified through whole-genome analysis of cultured isolates (6). Core genes, genes present in all genomes of *Prochlorococcus* sequenced, are shown in blue. Flexible genes, genes not present in all genomes of *Prochlorococcus* sequenced, are shown in pink.

Acknowledgements and author contributions

J.F.-L. and Y.S. contributed equally to this work. J.F.-L., Y.S., G.W.T., S.W.C., and E.F.D. designed research; Y.S. performed research; S.C.S. contributed new reagents/analytic tools; J.F.-L., Y.S., G.W.T., M.L.C., S.W.C., and E.F.D. analyzed data; and J.F.-L., Y.S., G.W.T., M.L.C., S.W.C., and E.F.D. wrote the paper. We thank the HOT team, the captain and crew of the R/V Kilo Moana for the expert assistance at sea, and Chon Martinez for preparing the sample DNA. This work was supported by the Gordon and Betty Moore Foundation (E.F.D. and S.W.C.), the National Science Foundation (S.W.C.), National Science Foundation Microbial Observatory Award MCB-0348001 (to E.F.D.), the Department of Energy Genomics GTL Program (E.F.D. and S.W.C.), and the Department of Energy Microbial Genomics Program (E.F.D. and S.W.C.). This article is a contribution from the National Science Foundation Science and Technology Center for Microbial Oceanography: Research and Education (C-MORE).

Supplementary Information for Chapter 2

Supplementary Methods
Supplementary Tables S1-S5
Supplementary Figures S1-S10

Supplementary Methods

Sample Collection for DNA Extraction

Bacterioplankton samples for DNA extraction were collected as previously described with minor modifications (Coleman et al., 2006). Briefly, the seawater was prefiltered in line through 125-mm Whatman GF/A filter (Whatman, Maidstone, U.K.) before the final collection of bacterioplankton cells onto 0.22-mm Steripak-GP20 filter (Millipore, Bedford, MA) using a Masterflex peristaltic pump (Cole Parmer Instrument Company, Vernon Hills, IL). After a total of 260 liters of seawater was filtered, the Steripak filter was covered with lysis buffer (50 mM Tris•HCl, 40 mM EDTA, and 0.75 M sucrose) and frozen in -80°C aboard before shipped frozen to the laboratory where they were stored at -80°C until DNA extraction.

DNA Extraction

DNA was extracted using slightly modified lysis and purification methods (Suzuki et al., 2004). Briefly, a solution of 5 mg/ml of lysozyme in 3 ml of lysis buffer was added to the Steripak-GP20 filter cartridge (Fisher, Fairlawn, NJ) after thawing, and incubated at 37°C for 30 min. Proteinase K (Sigma, St Louis, MO) in sterile water was added (at a final concentration of 0.5 $mg \times ml^{-1}$) into the Steripak-GP20 filter cartridge, followed by addition of SDS (Sigma, St Louis, MO) to a final concentration of 1%. The filter cartridges were sealed and incubated at 55°C for 20 min, followed by further incubation at 70°C for 5 min to further promote cell lysis. The lysate was remove from the filter cartridge, and nucleic acids were extracted twice with phenol:chloroform:IAA (25:24:1; Sigma, St Louis, MO) and once with chloroform:isoamyl

alcohol (24:1; Sigma). The purified aqueous phase was concentrated by spin dialysis using a Centricon 100 filter. An aliquot (~2 mg) of the extracted DNA was used for GS20 pyrosequencing.

Sample Collection for RNA Extraction

Bacterioplankton cells for total RNA extraction were collected filtering seawater from the same water sample that was used in DNA sample collection. We modified the collection process to shorten sampling time and improve sample preservation, which is critical in transcriptomics studies. The Niskin bottle transportation time in the water column depends entirely on the depth the CTD reaches; however, immediately upon shipboard retrieval of the CTD, a smaller volume of seawater (~ 1 liter) was filtered as rapidly as possible. The time from the start of filtration to storage in RNA later was 12 min. Briefly, the seawater was prefiltered through 1.6-mm GF/A filters (Whatman, Maidstone, U.K.) and then filtered through 25-mm and 0.22-mm Durapore filters (Millipore, Bedford, MA) using a four-head peristaltic pump system. The prefiltering step was used to remove most eukaryotic cells, although picoeukaryote cells (eukaryotes <2.0 mm in diameter) were present in the sample. The four Durapore filters (identica replicates) were immediately transferred to a screw-cap tube containing 1 ml of RNAlater (Ambion Inc., Austin, TX) after filtration, and frozen and kept at -80°C aboard the R/V Kilo Moana. Samples were transported frozen to the laboratory in a dry shipper and stored at -80°C until RNA extraction procedures.

RNA Extraction

Total RNA was extracted using a mirVana RNA isolation kit (Ambion, Austin, TX), with several modifications to recover RNA possibly released to the 1 ml of RNAlater due to the sample freeze and thaw. Samples were thawed on ice, and the 1 ml of RNAlater was gently pipetted out and loaded onto two Microcon YM-50 columns (Millipore, Bedford, MA) for desalting and concentrating by centrifugal filtration. The resulting 50 ml of RNAlater was added back to the sample tubes, and total RNA extraction was proceeded following the mirVana manual. Genomic DNA was removed using a Turbo DNA-free kit (Ambion, Austin, TX).

Finally, extracted RNA (DNase-treated) from four replicate filters were combined, purified, and concentrated by using the MinElute PCR Purification Kit (Qiagen, Valencia, CA).

Microarray Analysis of *Prochlorococcus* Gene Expression

For the experiments with Prochlorococcus MED4, cells were grown in the Pro-99 seawater-based medium (Rippka et al., 2000) at 21°C under continuous white light at 16 mol photon$\times m^{-1} \times s^{-1}$. Cells were harvested by centrifugation (10,000 ′ g) in log phase growth. Growth conditions and cell collection under phosphorus starvation of Prochlorococcus MIT 9313 were as described by Martiny et al. (Martiny et al., 2006). Samples of Prochlorococcus MIT 9313 were taken after 12 h under phosphorus starvation.

Before microarray analysis and RNA amplification, DNA was removed using the Turbo DNA-free kit (Ambion, Austin, TX). Synthesis, labeling, and hybridization of cDNA onto customized MD4-9313 Affymetrix (Santa Clara, CA) microarrays were performed following the standard Affymetrix protocol, and scanning was carried out according to Affymetrix protocols for Escherichia coli (www.affymetrix.com/support/technical/manual/expression_manual.affx). Data visualization was carried out by using GeneSpring software (version 7.3.1; Silicon Genetics, Palo Alto, CA). An initial normalization was applied using the Robust Multichip Average algorithm (Bolstad, Irizarry, Astrand & Speed, 2003) implemented in GeneSpring. Those values were later normalized using the lowess correction performed by using the software R (www.R-project.org).

RT-qPCR Analysis

Possible traces of DNA were removed using Ambion's Turbo DNA-free kit (Ambion, Austin, TX) following the manufacturer's instructions with minor modifications. The volume of Turbo DNase I was increased to 3 ml of Turbo DNase I (Ambion's Turbo DNA-free, Ambion) and the reaction mixture was incubated at 37°C for 60 min. RNA (1 ng) was reverse-transcribed with random hexamer primers and Superscript II reverse transcriptase (Invitrogen, Carlsbad, CA) following the manufacturer's instructions. Reverse transcription was performed at 42°C for 2 h, after an initial incubation step of 10 min at 25°C. The synthesized cDNA and purified

environmental DNA (1 ng) were used in SYBR green quantitative PCR (qPCR) using the specific primers for the genes of interest (Table S1). To compare the relative expression of genes we modified the $2^{-\Delta\Delta C_T}$ method (Livak & Schmittgen, 2001) and used the formula $cDNA\ DNA = (1+E_{DNA})^{C_T\ DNA} / (1+E_{cDNA})^{C_T\ cDNA}$ to take into consideration the different amplification efficiencies in separate qPCR runs.

Sequence Analyses of cDNA and DNA Reads

The defined bit score cutoff for assigning reads to GOS peptides and NCBI-nr protein was based on in silico tests using BLASTX comparisons against nonmarine microbial genomes (Figure S5) where a bit score of >40 was shown to result in low false positive frequencies (<2%). Furthermore, a breakdown of amino acid identity and length values for bit scores >40 observed in DNA library (Figure S5) highlights the stringency of this cutoff.

Assignment of reads to GOS protein clusters enabled the calculation of cluster-based expression ratio, a normalized comparison of the number of reads found for each protein cluster in the cDNA library relative to that found in the DNA library. To normalize this ratio for the difference in DNA and cDNA library size, the number of reads assigned to any given protein cluster was divided by the total number of reads in the respective library. The resulting cluster fraction for the cDNA library then was expressed as a function of the representation in DNA library. The cluster-based expression ratios were ranked from highest to lowest (Figure 1) to look at clusters being expressed at elevated levels.

The relative abundance of detected clusters was taken into consideration by dividing cluster-based expression ratios into categories based on their abundance in the DNA library. Using an empirical cumulative density function (Figure S8), clusters were categorized as low (<9 read members), medium (9-161 read members), high (161-461 read members), or extremely high abundance (>461 read members). This abundance measure also reflects the conservation of protein clusters, because more conserved proteins clusters are likely to have more members (e.g., RNA polymerase). Rarefaction analysis for each sample was based on best matches against the GOS database. The frequency of observed best matches to GOS protein clusters for each library

was used to calculate rarefaction curves with the program Analytic Rarefaction 1.3.

Putative Prochlorococcus reads were identified as reads with top BLASTX hit (against NCBI-nr) to Prochlorococcus and with a bit score >40. Each of these putative Prochlorococcus reads then was searched against a database of 11 whole-genome sequences using BLASTN and assigned to the best hit gene. For comparison with a single-reference genome, MIT9301, the assigned genes from 11 strains all were translated to their MIT9301 ortholog (Kettler et al., 2007), where one exists. The number of raw cDNA reads per gene was used to indicate the most transcribed genes in the entire Prochlorococcus population. To normalize cDNA reads per gene copy, the number of DNA reads per gene first was divided by the gene length (′1,000 to give reads per kb) to account for a clear direct relationship between gene length and its representation in the DNA reads (Figure S9). A clear, direct relationship with gene length does not exist for cDNA reads. The number of cDNA reads per gene then was divided by this normalized DNA (DNA reads per kb) to give an indication of per-copy cDNA abundance. This additional normalization to gene length, which is not possible for the whole community without good reference genomes, is generally consistent with the expression ratio (cDNA/DNA)-analogous to the cluster-based expression ratio used for whole-community analyses-except, for example, in cases of very short genes (Figure S9).

Removal of Low-Quality and Ribosomal RNA (rRNA) GS20 cDNA Sequences. Polymeric sequences inadvertently introduced into the cDNA library during cDNA synthesis (via polyadenylation of mRNA/aRNA and subsequent amplification step) were trimmed from reads based on the observed frequency of polymeric sequences in the DNA library (Figure S10). A noticeable peak in poly(A/T) sequences in the cDNA library around 16 bp (Figure S10) is attributable to polyadenylation of the mRNA and subsequent amplification with a T7-BmpI-$(dT)_{16}$VN primer. To remove residual T7 promoter and priming sites not cleaved by BmpI, reads were initially screened by using cross-match (-minmatch 10, -minscore 10; found in 32,246 reads). Reads containing a poly(A/T) sequence >10 bp (cutoff based on Figure S10) or multiple poly(A/T) runs in a single read (4 ′ 6 bp) were trimmed unless a significant BLASTN match across the polymeric sequence in the cDNA read was identified in a read from the DNA library

(39,444 reads remained untrimmed). By using these criteria, bases flanking the ends of each cDNA read were trimmed, and reads with polymeric sequences located in the middle of reads were deemed putative chimeras and removed from the dataset (5,232 chimeric reads).

rRNAs were removed from the cDNA library by using a combined 5S, 16S, 18S, 23S, and 28S rRNA database derived from available microbial genomes and sequences from the ARB SILVA LSU and SSU databases (www.arb-silva.de). BLASTN matches with bit score >40 were considered significant and deemed rRNA sequences (65,859 reads; 51.3% of reads). This bit score cutoff resulted in <1.7% false positives against a database of all non-rRNA microbial genes from available microbial genomes. After trimming and removal of rRNAs, 54,568 reads (average length 95 bp) totaling 5,194,332 bp remained in the cDNA sample. Raw metagenomic GS20 DNA and cDNA reads have been deposited in GenBank.

MEGAN and Statistical Analysis

We performed sequence comparisons of DNA and cDNA pyrosequencing results against the NCBI-nr database. Only the best hit of the top BLASTX hits with a bit score >40 was used for MEGAN analysis (version 2beta3, August 2007). MEGAN is a new software program (Huson et al., 2007) used to explore the taxonomical content of the dataset, employing the NCBI taxonomy to summarize and order the results. Moreover, MEGAN gives the number of hits obtained for the different taxonomic groups, which allows for statistical comparison of the distribution of those groups on the phylogenetic trees. Statistical differences between taxonomic groups on the DNA and cDNA trees obtained in MEGAN was assessed using the software R (www.R-project.org). c^2 test was used to estimate differences at the level of kingdom. In this case, we used the Pearson's c^2 test with simulated P value (based on 10,000 replicates) and the log likelihood ratio (G test) test with Williams' correction (g.test.r code in R, from Peter L. Hurd, www.psych.ualberta.ca/~phurd/cruft/).

Supplementary Tables and Figures

Table S1. Oligonucleotide used for qPCR analysis of genes identified by pyroseqeuncing. Sequences were compared against the NCBI-nt database of nucleotide sequences using BLASTn.

Best hit in nr database	Oligonucleotide sequences 5'-3'	Comments
Common, highly expressed		
Thioredoxin peroxidase (Tpx)	TAT TAA GTG CTG AGA AAT CTT GA TGG GTT GTT CTA TTC TTT TAC CC	Specific only for *Prochlorococcus* MIT9312
Ammonium transporter (Amt)	ATTGGATTTGGAATTATGTATTAC AGTATTCCAGGAATTATTTCC	Specific only for *Prochlorococcus* MIT9312
Photosystem I PsaL protein (subunit XI) (PsaL)	TTG TTA ATC CGC CAA AGG AC AAG CAA AAA CAG CTC CTC CA	Amplifies *Prochlorococcus* MIT9301 and AS9601
Common, low expressed		
Alanyl-tRNA synthetase (AlaRS)	CAG ACA TGG GAG ATT TGT TAG G TCA GGA TAA TTA TTT TGC ATT AAA	Amplifies *Prochlorococcus* MIT9312 and MIT9301
Transcription-repair coupling factor (TRCF)	AAG GTT GAA ATC TAT TAT TTA TTG TTC TTA CAT CAG GCA AAC AGG TAA	Amplifies *Prochlorococcus* MIT9312, MIT9301 and AS9601
Phosphoribosylformylglycin amidine synthase II (FGAM synthaseII)	GCAGCAATAGTTCCTCTAAAAGGG TTC TGG TGT TGC TGC TTC TG	Amplifies *Prochlorococcus* MIT9312 and MIT9515
Cobaltochelatase, CobN subunit (CobN)	TTTTAATGCGAATGCTATTTGCC CCT ATA GAT TTG CCA GGT AAC CA	Amplifies *Prochlorococcus* MIT9301,MIT 9515, AS9601, MIT 9312 and MED4
Cobyrinic acid a,c-diamide synthase (CbiA)	AAG AGA ATT CAT ATT TCA AAG AAT GTT CCA ACC TAT TTG CAG GAA TTT	Amplifies *Prochlorococcus* 9301, 9515, AS9601, 9312 and MED4
Only in cDNA library		
Putative light-harvesting protein alpha chain (LHC)	AGCAATGATACATCTTGTTCTGC AGT TGC TGC TGC CTC AAA C	Specific for uncultured proteobacterium eBACred25D05
Predicted xylene monooxygenase hydroxylase component (XylM)	TTTGCAGTGTGATAACTCAT TGTGCTATCAACAGGTATATTTGCCGG	Specific for uncultured bacterium BAC13K9BAC

Table S2. Representatives of the GOS protein clusters that are unique to 75-m cDNA library.

Cluster ID	Abundance	GO term	Pfam	TIGRfam	NR
14275698	667	–	–	–	–
11297554	28	–	–	–	ZP_01470602.1| hypothetical protein RS9916_32857 [Synechococcus sp. RS9916]
14230436	19	–	–	–	AAT90307.1| putative light-harvesting protein alpha chain [uncultured proteobacterium eBACred25D05]
12073604	14	photosynthesis\light reaction	–	–	ZP_01583951.1| antenna complex, alpha/beta subunit [Dinoroseobacter shibae DFL 12]
12023158	8	–	–	–	ZP_01470602.1| hypothetical protein RS9916_32857 [Synechococcus sp. RS9916]
11899146	6	–	–	–	YP_001008748.1| hypothetical protein A9601_03531 [Prochlorococcus marinus str. AS9601]
7478	4	translational initiation	–	–	–
11393514	4	photosynthesis\light reaction	–	–	AAT90308.1|putative light-harvesting protein beta chain [uncultured proteobacterium eBACred25D05]
19661	3	–	–	–	putative proteorhodopsin [uncultured bacterium]
11054015	3	–	–	–	CAL01029.1| chlorophyll a/b binding light harvesting protein pcbA [uncultured Prochlorococcus sp.]
16914	3	–	–	–	ZP_01255953.1| Substrate-binding region of ABC-type glycine betaine transport system [Psychroflexus torquis ATCC 700755]
17232	2	transcription	–	–	EAZ99485.1|DNA-directed RNA polymerase subunit beta [Marinobacter sp. ELB17]
14025838	2	transport	–	–	ZP_00949339.1|putative outer membrane protein [Croceibacter atlanticus HTCC2559]
14212924	1	–	TonB-dependent receptor	–	–

Table S3. Representatives of the GOS protein clusters that are unique to 75-m DNA library.

Cluster ID	Abundance	GO term	Pfam	TIGRfam	NR
174	333	de novo' IMP biosynthesis	–	–	GAR transformylase 2 [Prochlorococcus marinus str. MIT 9301]
260	245	–	–	–	Glycosyl transferase, family 2 [Prochlorococcus marinus str. MIT 9301]
5431	241	lipopolysaccharide biosynthesis	–	–	Glycosyl transferase, family 2 [Prochlorococcus marinus str. MIT 9301]
700	209	mismatch repair	–	–	putative DNA mismatch repair protein MutS family [Prochlorococcus marinus str. AS9601]
442	200	–	–	small_GTP: small GTP-binding protein domain	Small GTP-binding protein domain [Prochlorococcus marinus str. MIT 9312]
3200	198	urea metabolism	Amidohydrolase family	urease_alph: urease, alpha subunit	Urease alpha subunit [Prochlorococcus marinus str. MIT 9301]
3868	196	lipopolysaccharide biosynthesis	–	–	UDP-N-acetylglucosamine pyrophosphorylase [Prochlorococcus marinus str. MIT 9301]
152	193	cobalamin biosynthesis	–	–	precorrin-2 C20-methyltransferase [uncultured Prochlorococcus marinus clone ASNC2259]
428	190	tryptophanyl-tRNA aminoacylation	–	trpS: tryptophanyl-tRNA synthetase	Tryptophanyl-tRNA synthetase [Prochlorococcus marinus str. AS9601]
1225	190	coenzyme A biosynthesis	–	–	ATP/GTP-binding site motif A (P-loop) [Prochlorococcus marinus str. AS9601]
4133	184	amino acid biosynthesis	Homoserine dehydrogenase	–	YP_001009547.1| Homoserine dehydrogenase:ACT domain-containing protein [Prochlorococcus marinus str. AS9601]
2731	180	intracellular protein transport	–	chloroplast envelope protein translocase, IAP75 family	outer envelope membrane protein-like protein [Prochlorococcus marinus str. AS9601]
3940	176	GTP biosynthesis	Radical SAM superfamily		Fe-S oxidoreductase [Prochlorococcus marinus str. MIT 9301]
288	173	–	–	–	DEAD/DEAH box helicase:Helicase C-terminal domain-containing protein [Prochlorococcus marinus str.AS9601]
133	172	pentose-phosphate shunt	Transaldolase	transaldolase	Transaldolase [Prochlorococcus marinus str. AS9601]
2871	171	electron transport	Pyridine nucleotide-disulphide oxidoreductase		Selenide,water dikinase [Prochlorococcus marinus str. MIT 9301]

Table S4. Taxonomic diversity of DNA and cDNA libraries computed by MEGAN after removal of rRNA sequences from the databases. BLASTx results with a bits-score cutoff of 40 were used to construct the trees. Color-coding corresponds to that in Figure 2. Bacteria: green; archaea: red; eukaryota: blue; viruses: light blue. Taxa within each kingdom have been ordered by rand abundance based on the total number of hits in the DNA library.

Phylum	Number of hits in the DNA library	Number of hits in the cDNA library	Percentage (%) of hits in the DNA library	Percentage (%) of hits in the cDNA library
Cyanobacteria	142,084	4,167	34.313	7.636
Proteobacteria	50,506	2,413	12.197	4.422
Bacteroidetes	9,943	375	2.401	0.687
Firmicutes	2,477	243	0.598	0.445
Actinobacteria	1,507	26	0.364	0.048
Planctomycetes	561	8	0.135	0.015
Chlorobi	517	9	0.125	0.016
Chloroflexi	335	13	0.081	0.024
Spirochaetes	251	6	0.061	0.011
Acidobacteria	219	5	0.053	0.009
Thermotogae	191	0	0.046	0
Deinococcus-Thermus	113	0	0.027	0
Verrucomicrobia	112	0	0.027	0
Fusobacteria	83	0	0.020	0
Aquificae	63	0	0.015	0
Chlamidiae	47	2	0.011	0.004
Nitrospirae	41	0	0.010	0
candidate division WS3	7	0	0.002	0
Unclassified bacteria	4	0	0.001	0
Candidate division OP8	4	0	0.001	0
Candidatus Poribacteria	4	0	0.001	0
Dictyoglomi	2	0	0.0005	0
Euryarchaeota	708	10	0.171	0.018
Crenarchaeota	168	0	0.041	0.000
Nanoarchaeota	3	0	0.001	0.000
Streptophyta	509	18	0.123	0.033
Chordata	495	21	0.120	0.038
Ascomycota	468	4	0.113	0.007
Chlorophita	307	9	0.074	0.016
Arthropoda	257	15	0.062	0.027
Ciliophora	167	16	0.040	0.029
Apicomplexa	166	6	0.040	0.011
Cnidaria	157	0	0.038	0.000
Mycetozoa	140	13	0.034	0.024
Echinodermata	110	3	0.027	0.005

Table S5. Top 20 *Prochlorococcus* highly expressed genes in the cDNA library depending on the kind of normalization applied on the dataset.

Raw cDNA		cDNA/DNA		cDNA/(DNA/length)	
P9301_0286*	Ammonium transporter family	P9301_11381	no description	P930*_02861	Ammonium transporter family
P9301_1715*	Photosystem I PsaB protein	P9301_03541	no description	P930*_17151	Photosystem I PsaB protein
P9301_1716*	Photosystem I PsaA protein	P9301_07111	no description	P930*_17161	Photosystem I PsaA protein
P9301_0576*	Ribulose bisphosphate carboxylase, large chain	P9301_04361	Predicted protein	P930*_03541	no description
P9301_0654*	Chlorophyll a/b binding light harvesting protein PcbD	P9301_03421	Photosystem II reaction center M protein (PsbM)	P930*_05761	Ribulose bisphosphate carboxylase, large chain
P9301_1138*	no description	P9301_13581	no description	P930*_03401	Photosystem II PsbB protein (CP47)
P9301_0340*	Photosystem II PsbB protein (CP47)	P9301_02911	conserved hypothetical protein	P930*_11381	no description
P9301_1699*	Elongation factor Tu	P9301_02861	Ammonium transporter family	P930*_16991	Elongation factor Tu
P9301_1350*	Photosystem II PsbC protein (CP43)	P9301_00661	Predicted protein	P930*_13501	Photosystem II PsbC protein (CP43)
P9301_1392*	polyribonucleotide nucleotidyltransferase	P9301_17021	30S ribosomal protein S12	P930*_13921	polyribonucleotide nucleotidyltransferase (pnp)
P9301_1349*	Photosystem II PsbD protein (D2)	P9301_09571	50S ribosomal protein L28	P930*_16741	RNA polymerase beta prime subunit
P9301_1674*	RNA polymerase beta prime subunit	P9301_12251	Possible high light inducible protein	P930*_07111	no description
P9301_0429*	30S ribosomal protein S4	P9301_05771	Ribulose bisphosphate carboxylase, small chain	P930*_17001	Elongation factor G
P9301_1226*	Porin homolog	P9301_02221	50S ribosomal protein L10	P930*_13491	Photosystem II PsbD protein (D2)
P9301_0711*	no description	P9301_17121	photosystem I subunit VIII (PsaI)	P930*_10121	thioredoxin peroxidase
P9301_0291*	conserved hypothetical protein	P9301_10121	thioredoxin peroxidase	P930*_04291	30S ribosomal protein S4
P9301_1700*	Elongation factor G	P9301_04291	30S ribosomal protein S4	P930*_02221	50S ribosomal protein L10
P9301_1012*	thioredoxin peroxidase	P9301_16451	ATP synthase subunit c	P930*_02451	no description
P9301_1711*	Photosystem I PsaL protein (subunit XI)	P9301_0321*	Cytochrome b559 alpha-subunit	P930*_06541	Chlorophyll a/b binding light harvesting protein PcbD
P9301_1803*	Transketolase	P9301_06071	plastocyanin	P930*_18031	Transketolase

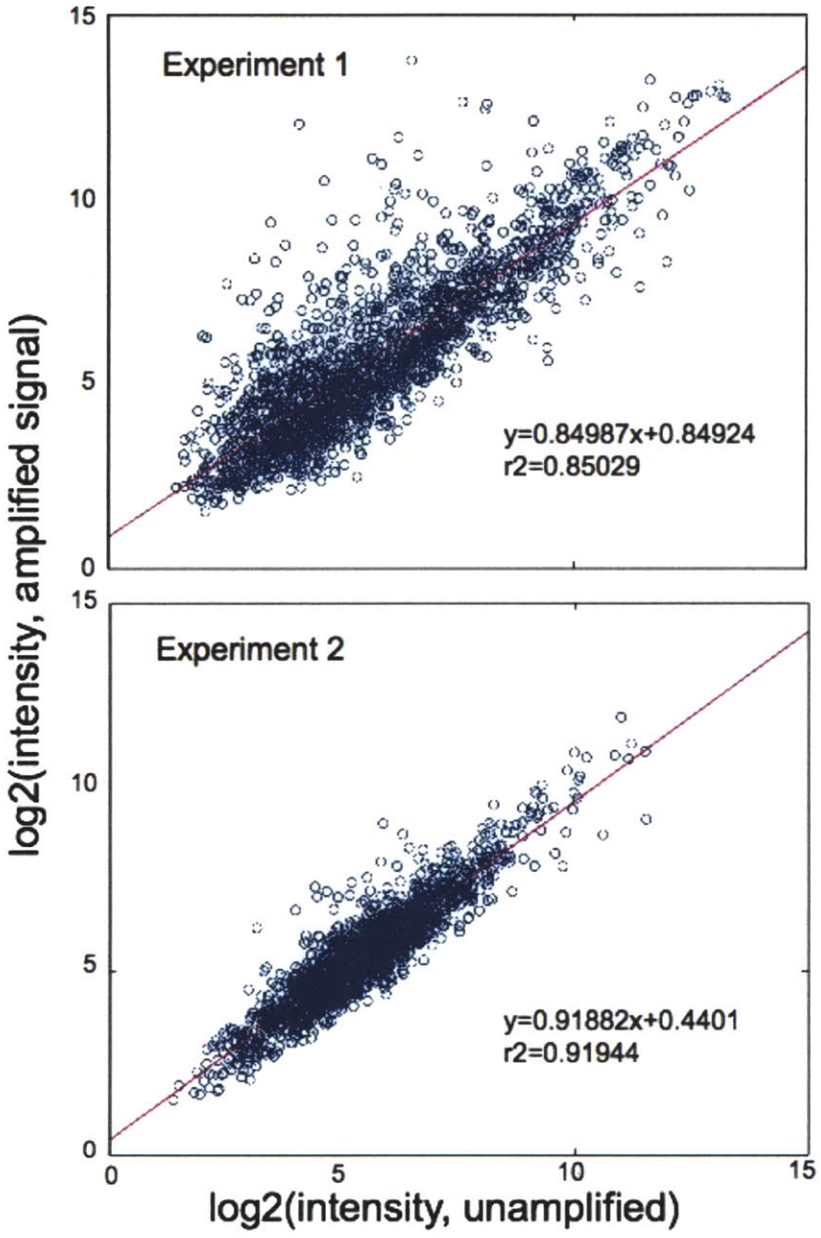


Figure S1. Comparison of linearly amplified and unamplified mRNA from cultures of *Prochlorococcus* (MED4) cells using custom Affymetrix arrays. Expression values for protein-coding genes of *Prochlorococcus* MED4 for unamplified RNA vs. the amplified RNA obtained from a 100-ng aliquot from the former. Results from two independent experiments are shown.

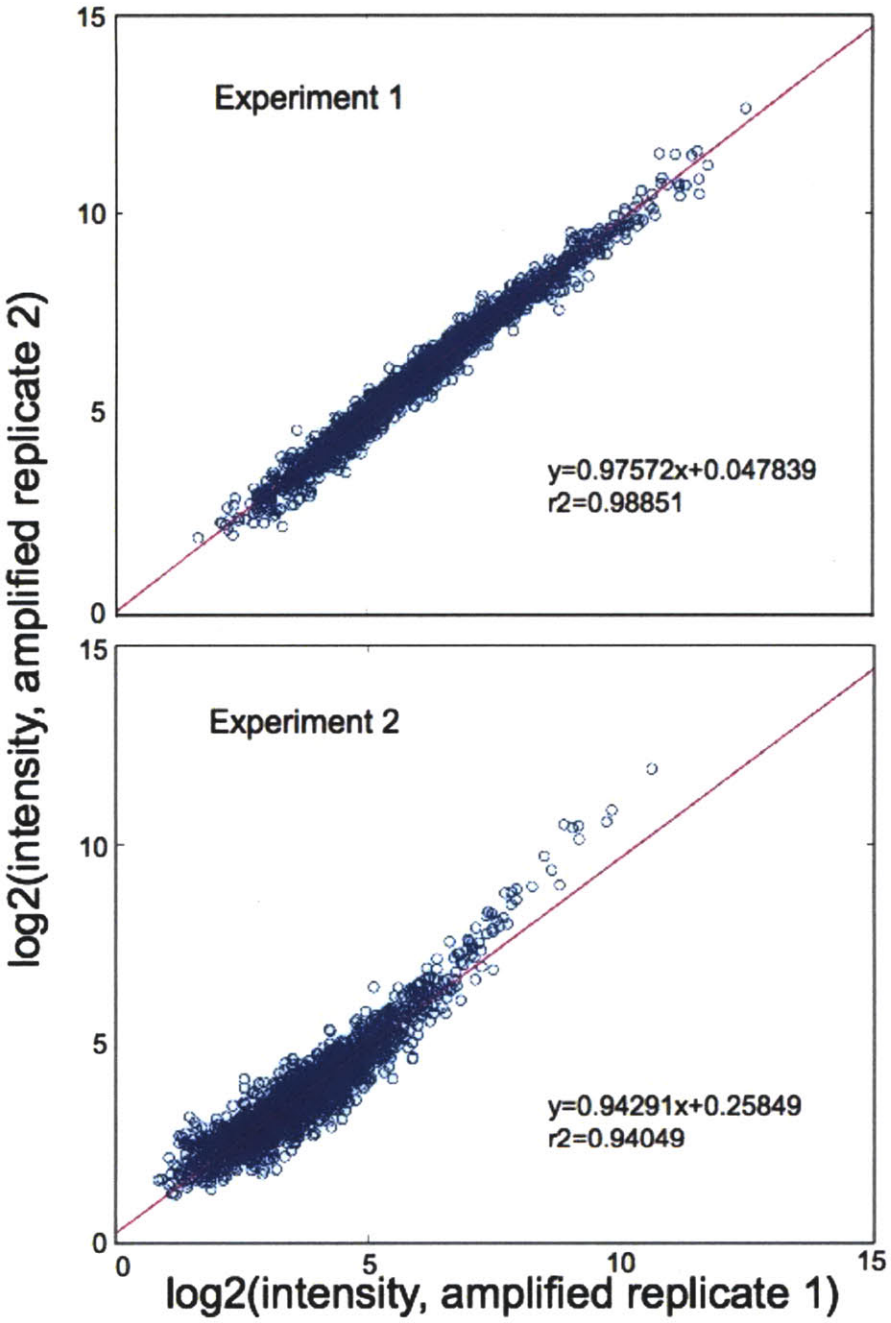


Figure S2. Comparison of linearly amplified mRNA from duplicate cultures of *Prochlorococcus* (MED4) cells using custom Affymetrix arrays. Expression values for protein-coding genes of *Prochlorococcus* MED4 of replicate amplified samples plotted against each other showing the reproducibility of the amplification. Results are from two independent experiments.

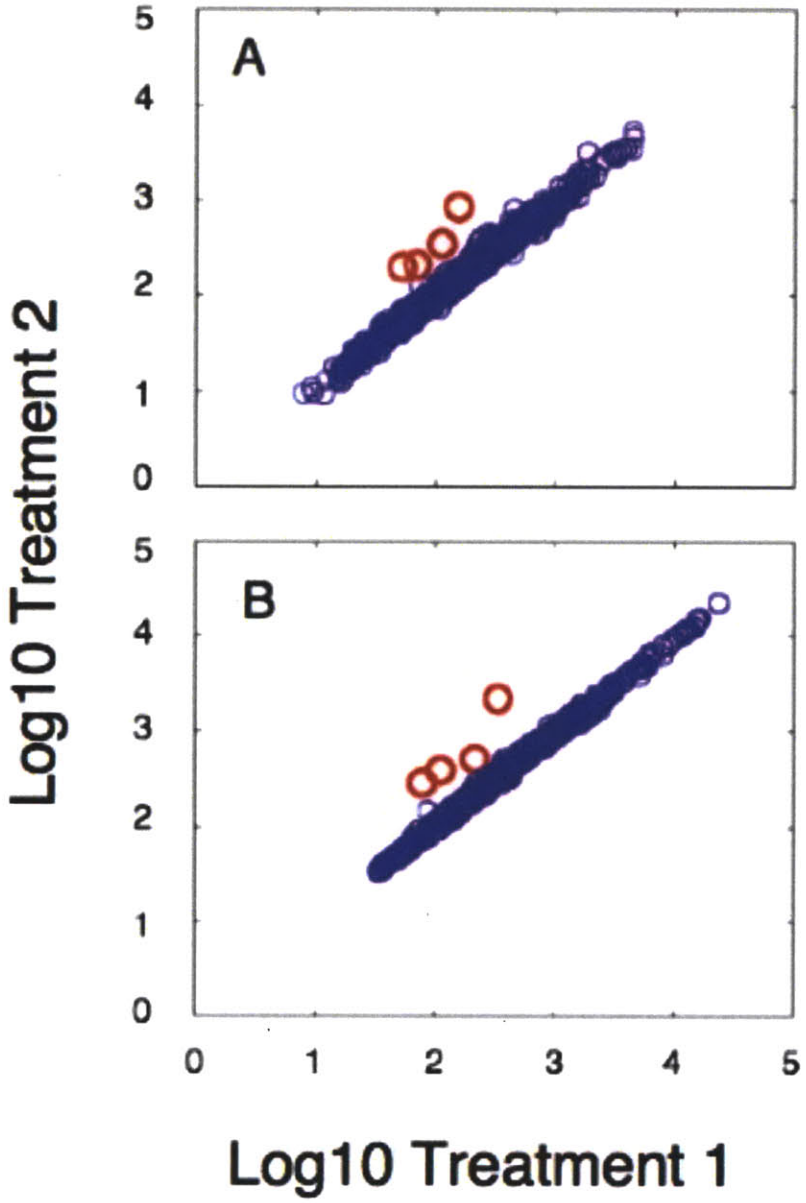


Figure S3. Comparison of the results of an experiment designed to reveal up-regulated genes in *Prochlorococcus* (MIT9313) under phosphate starvation, using unamplified (*A*) and amplified (*B*) RNA using custom Affymetrix arrays. Treatment 1: Control culture in phosphate-replete media. Treatment 2: phosphate-starved cultures. The same four genes appear as differentially expressed in both amplified and unamplified treatments: a phoB two component response regulator, a Som like protein (phosphate-limitation inducible outer membrane porins), and two ABC transporter substrate (phosphate) binding protein.

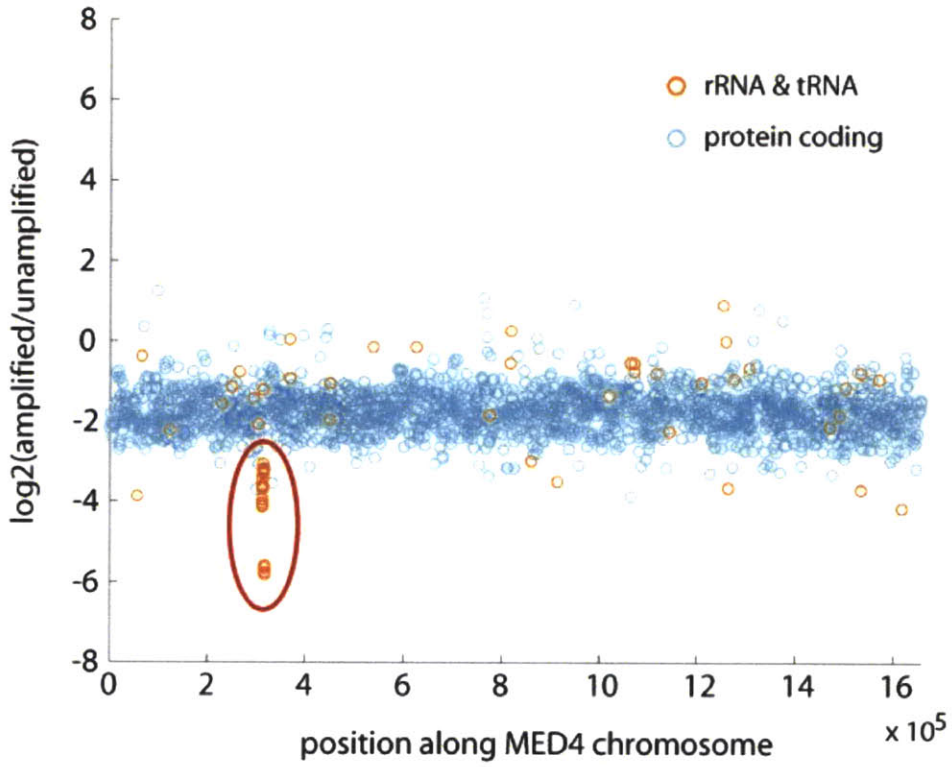


Figure S4. Analysis of accuracy of RNA amplification as a function of position along the *Prochlorococcus* MED4 chromosome using custom Affymetrix arrays. The ratio of the expression values yielded from amplified and unamplified RNA for protein-coding genes (blue) and ribosomal RNAs and tRNAs (red dots). The circled red dots are rRNAs.

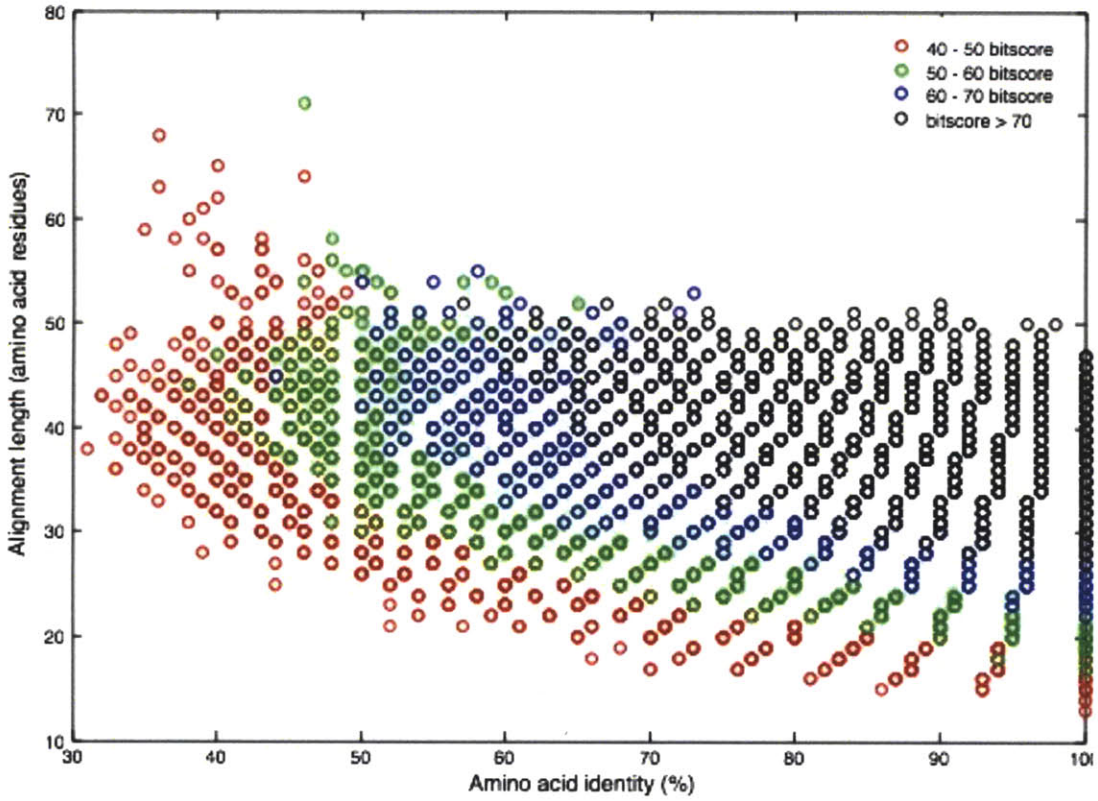


Figure S5. Stringency of the BLASTX bit score cutoff, in terms of alignment length and amino acid identity. Each circle represents an alignment between a cDNA pyrosequencing read and an NCBI-nr database sequence. Alignments with a bit score >40 were considered significant in our analyses.

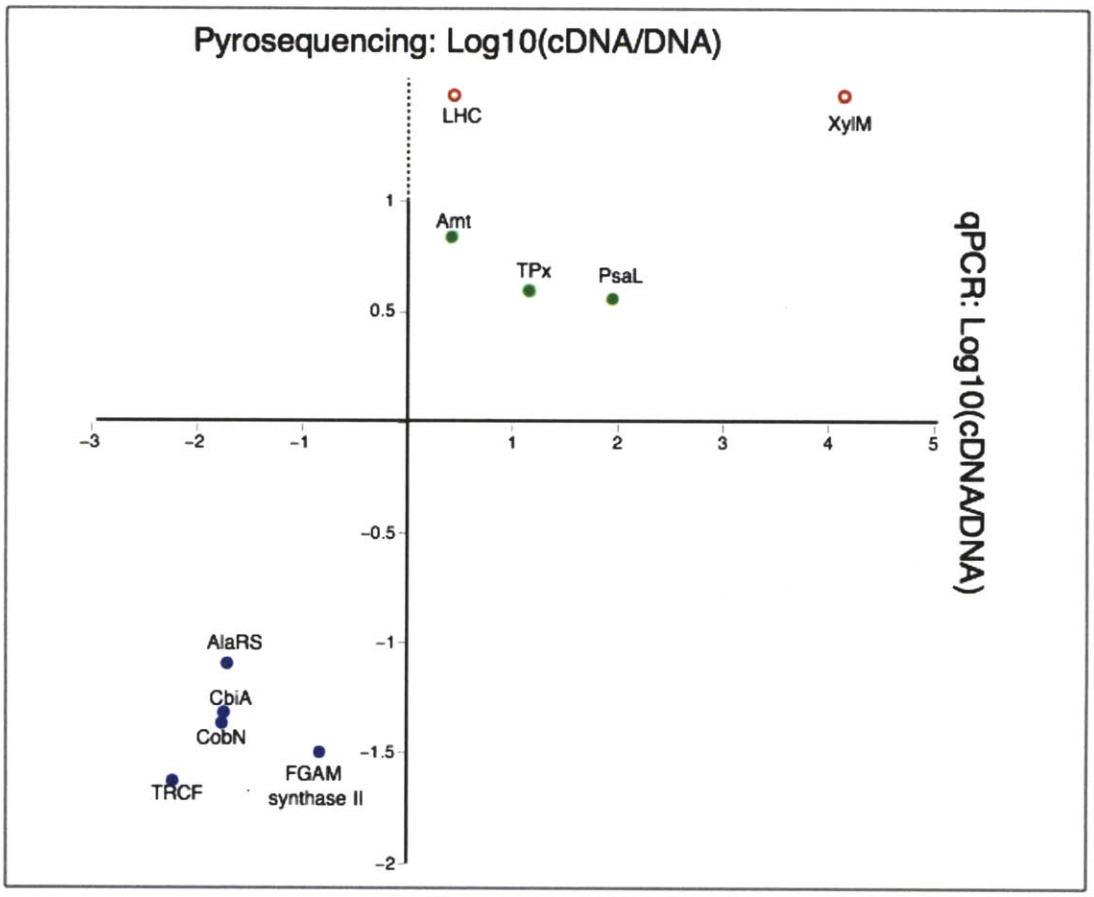


Figure S6. Comparison of transcriptional levels of selected genes using pyrosequencing and RT-qPCR/qPCR. The unamplified environmental RNA and DNA samples were used for quantitative PCR. The cDNA to DNA ratio in qPCR analysis (x axis) was calculated based on the modified $2^{-\Delta\Delta C_T}$ method (see Supplementary Methods). The cDNA to DNA ratio in pyrosequence analysis (y axis) was normalized to the size of the respective libraries. More specifically, the ratio was calculated as the fraction of reads assigned to the targeted gene in the cDNA library divided by that in the DNA library. Three sets of genes were selected based on their enrichment in the cDNA pyrosequence library. Green solid circle: genes with normalized cDNA/DNA ratio >1. Blue solid circle: genes with normalized cDNA/DNA ratio <1. Red open circle: gene only detected in the cDNA library but not in the DNA library, and thus the cDNA/DNA ratio could not be calculated for pyrosequencing data (dotted part of y axis). The full names of the 10 selected genes are listed in Table S1.

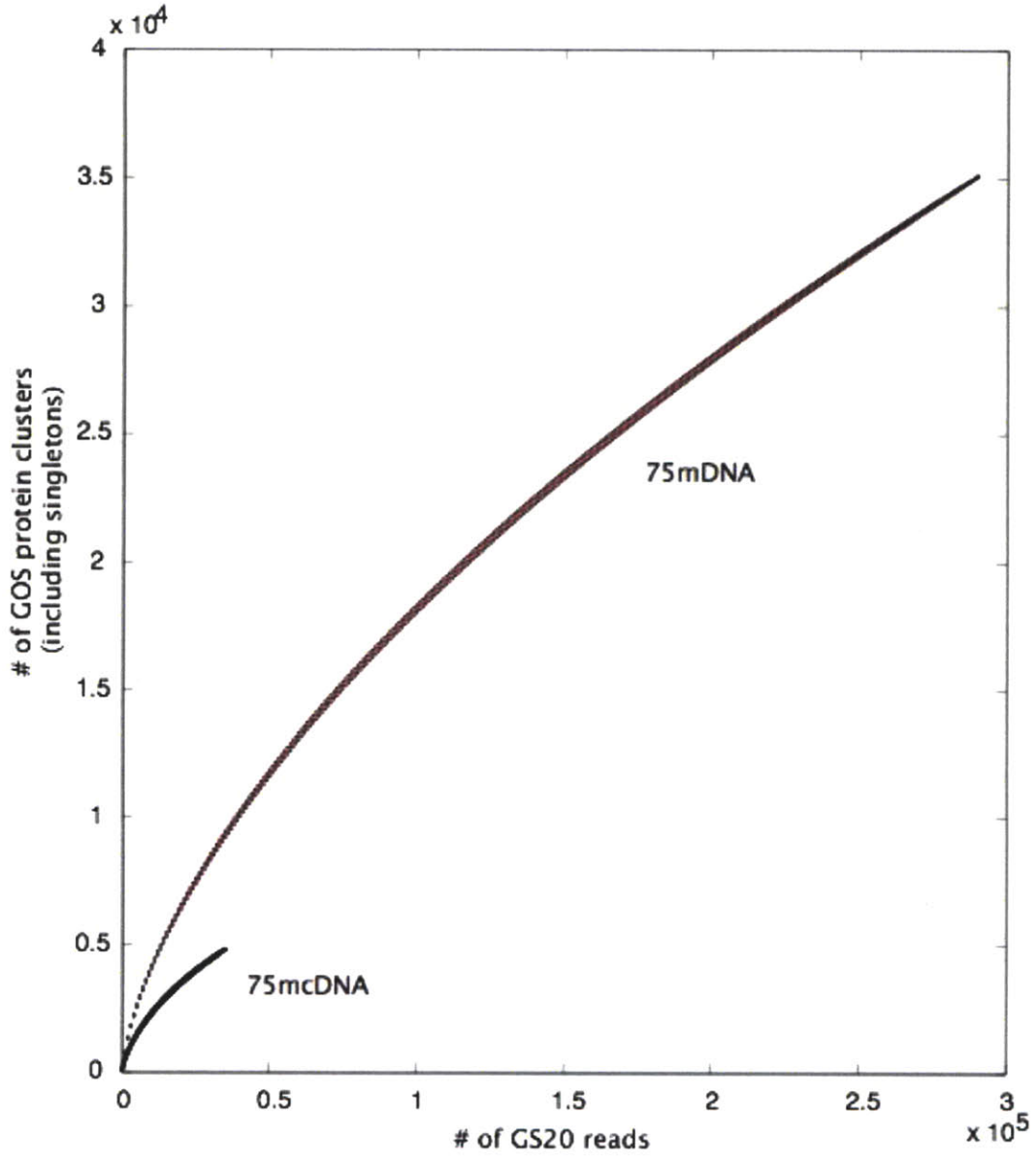


Figure S7. Rarefaction analyses for cDNA and DNA libraries. The rarefaction analysis was based on the frequency of significant BLASTX matches in the GOS peptide database, with increasing number of Roche GS20 DNA pyrosequencing reads. Red dots represent the average values, and the black dots represents the 95% confidence interval values.

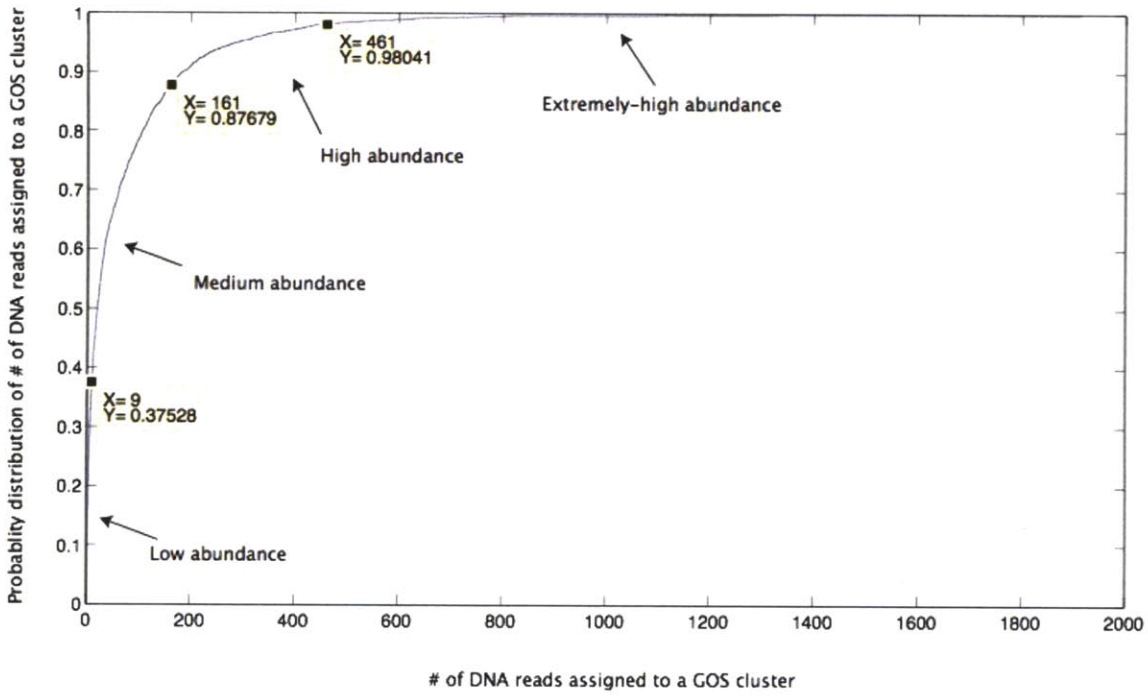


Figure S8. Empirical cumulative probability density function of the number of DNA reads assigned to a GOS protein cluster. The GOS protein clusters were arbitrarily binned to low, medium, high, and extremely high categories. Boundary values for each category, e.g., the number of DNA reads assigned to the cluster and its probability, also are shown.

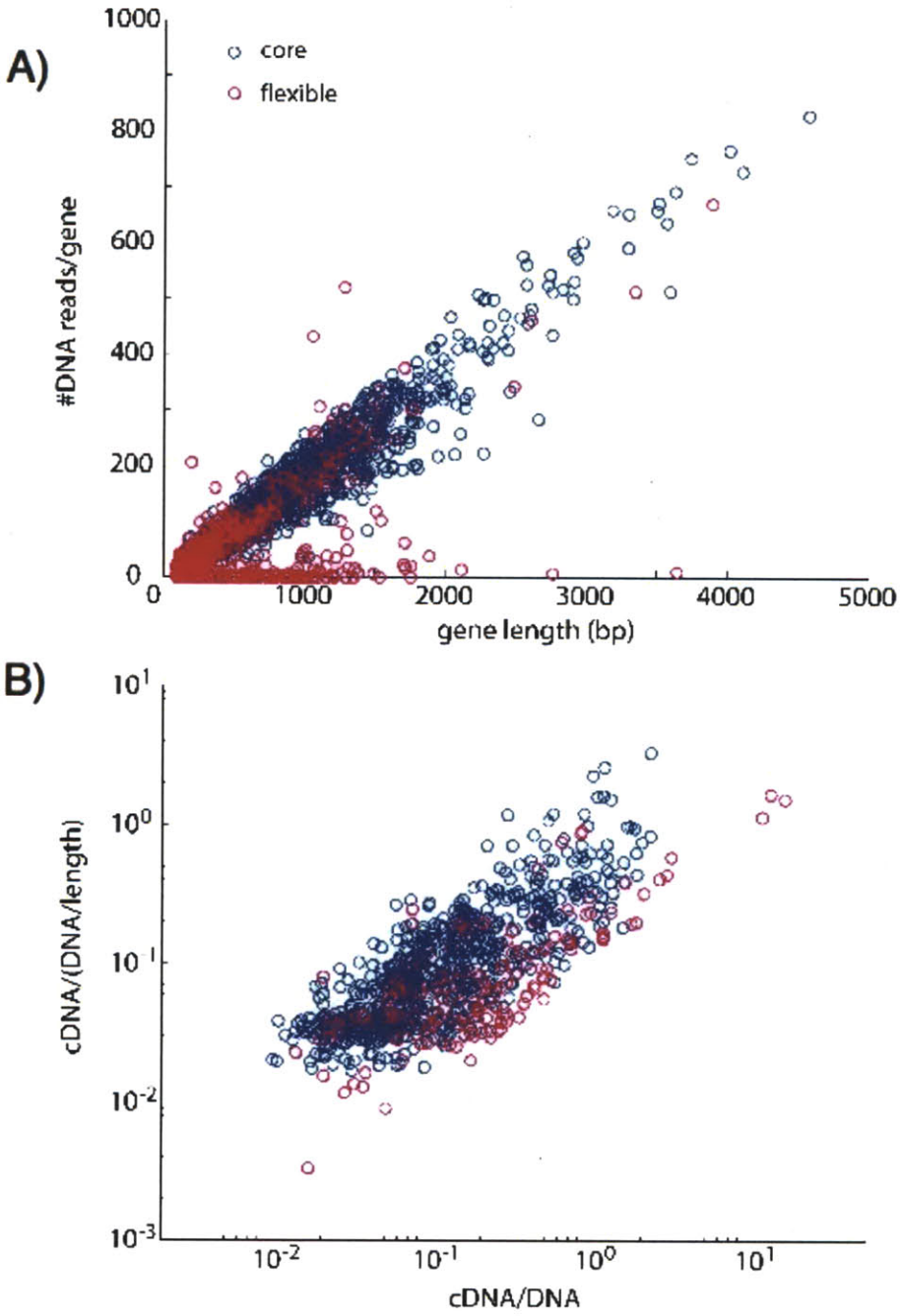


Figure S9. Effect of gene length on the number of hits in the DNA library, assessed by using *Prochlorococcus* MIT9301. (*A*) Linear relationship between the number of hits in the DNA database and gene length in the genome of MIT9301. (*B*) Relationship between the normalized cDNA against DNA hits and the normalized cDNA already normalized against gene length. In blue, core genes, i.e., genes present in all genomes of *Prochlorococcus* sequenced to date. In pink, flexible genes, i.e., genes not present in all genomes of *Prochlorococcus* sequenced.

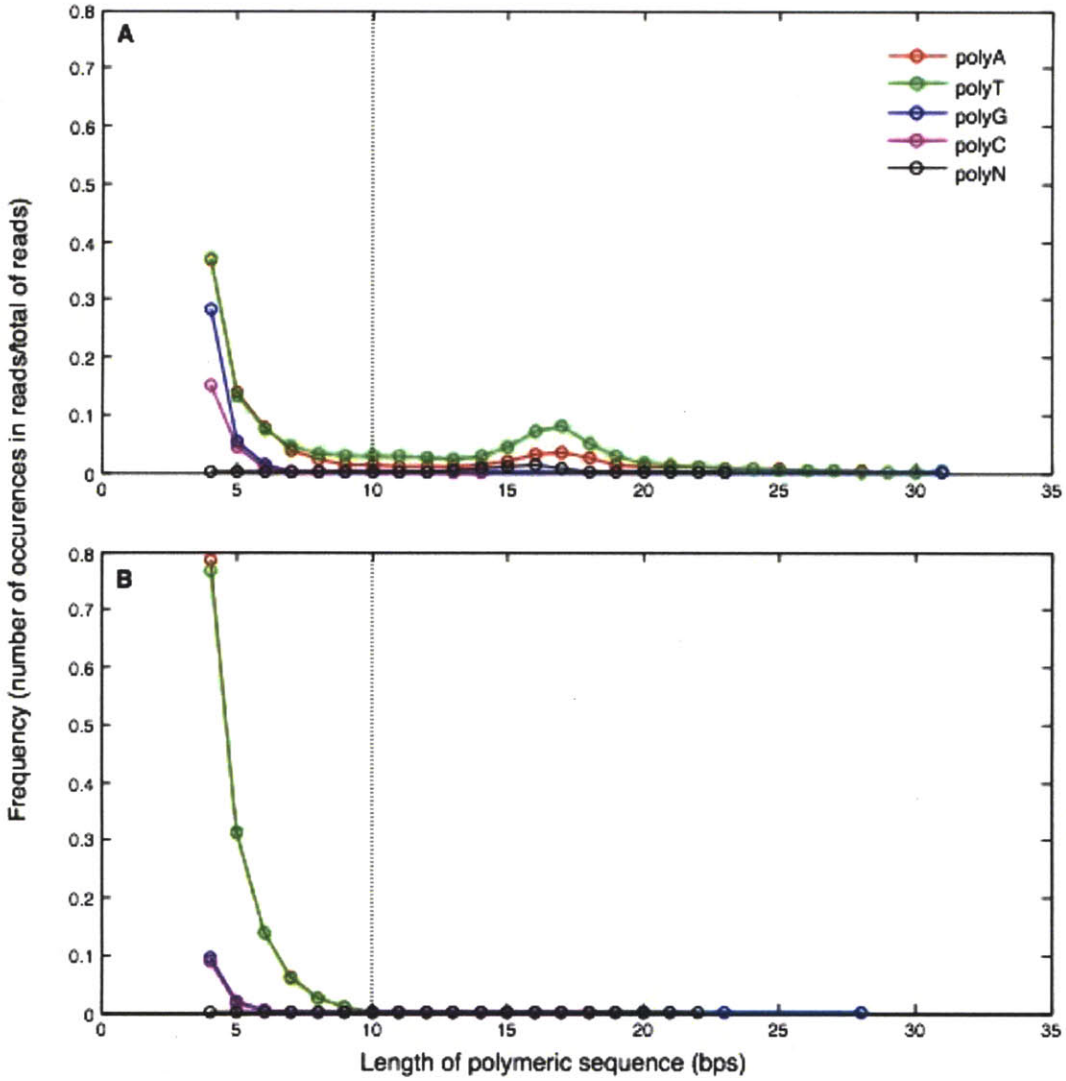


Figure S10. Distribution of the frequency of polymeric nucleotide sequence (A, T, G, C, and N) lengths found in the 75-m cDNA (*A*) and 75-m DNA (*B*) pyrosequencing libraries. The peak in polymeric sequence length at 15-16 bp in the cDNA reads is a result of the polyadenylation in library preparation. The dashed line at 10 bp indicates the cutoff used in the trimming of the cDNA data.

CHAPTER THREE

Integrated metatranscriptomic and metagenomic analyses of stratified microbial assemblages in the open ocean

Yanmei Shi, Gene W. Tyson, John M. Eppley, Edward F. DeLong

This chapter is presented with slight formatting modification, as it appeared in *The ISME Journal* advance online publication, 9 December 2010; doi:10.1038/ismej.2010.189. Corresponding supplementary information is appended.

Chapter 3: Integrated metatranscriptomic and metagenomic analyses of stratified microbial assemblages in the open ocean

Abstract

As part of an ongoing survey of microbial community gene expression in the ocean, we sequenced and compared ~38 Mbp of community transcriptomes and ~157 Mbp of community genomes from four bacterioplankton samples, along a defined depth profile at Station ALOHA in North Pacific subtropical gyre (NPSG). Taxonomic analysis suggested that the samples were dominated by three taxa: Prochlorales, Consistiales, and Cenarchaeales, that comprised 36-69% and 29-63% of the annotated sequences in the four DNA and four cDNA libraries, respectively. The relative abundance of these taxonomic groups was sometimes very different in the DNA and cDNA libraries, suggesting differential relative transcriptional activities per cell. For example, the 125m sample genomic library was dominated by *Pelagibacter* (~36% of sequence reads), which contributed far fewer sequences to the community transcriptome (~11%). Functional characterization of highly expressed genes revealed taxon-specific contributions to active biogeochemical processes. Examples included *Roseobacter*-relatives involved in aerobic anoxygenic phototrophy at 75m, and the unexpected contribution of low abundance crenarchaea to ammonia oxidation at 125m. Read recruitment using reference microbial genomes indicated depth-specific partition of coexisting microbial populations, as highlighted by the transcriptionally active HL-like *Prochlorococcus* population in the bottom of the photic zone. Additionally, nutrient uptake genes dominated *Pelagibacter* transcriptomes, with apparent enrichment for certain transporter types (e.g., the C4-dicarboxylate transport system) over others (e.g., phosphate transporters). In total, the data support the utility of coupled DNA and cDNA analyses for describing taxonomic and functional attributes of microbial communities in their natural habitats.

Introduction

Marine microbial communities, centrally involved in the fluxes of matter and energy in the global oceans, are major drivers of global biogeochemical cycling (Arrigo, 2005; Karl & Lukas, 1996). Our knowledge of abundance, diversity and gene content of planktonic microbes has been fundamentally advanced over the past three decades, by both model organism-based studies (Coleman & Chisholm, 2007; Giovannoni et al., 2005b), as well as metagenomic surveys of natural microbial communities (DeLong et al., 2006; Dinsdale et al., 2008; Rusch et al., 2007). In particular, metagenomic comparisons of distinct microbiomes (DeLong et al., 2006; Dinsdale et al., 2008) have revealed habitat-dependent distribution of taxons and gene families, likely shaped by the biogeochemical conditions of each environment. Clearly, determining if and how such genomic variations are manifested at the level of gene expression and regulation represents another critical step towards understanding the interplay between microbes and their

natural environment, as well as their metabolic strategies to exploit distinct ecological niches.

Metatranscriptomics involves the direct sampling and sequencing of gene transcripts from natural microbial assemblages, and provides quantitative assessment of microbial gene expression, without requiring *a priori* knowledge of community taxonomic and genomic compositions. We first carried out a pilot metatranscriptomic study at the Hawaii Ocean Time-series (HOT) Station ALOHA (Frias-Lopez et al., 2008), where community transcripts were analyzed in parallel with genomic sequences for a bacterioplankton assemblage at 75m depth (within the mixed layer). One unexpected finding from that study was that many highly abundant transcripts (most of which were designated as hypothetical genes) were absent or in low abundance in the coupled DNA library, suggesting they originated from low abundance microorganisms (or less frequently represented genes in hypervariable genomic regions). Subsequently, comparative analyses of surface water samples have shed light on the day/night and geographical differences in community gene expression (Hewson, Rachel, Tripp, Joseph & Jonathan, 2010; Poretsky et al., 2009). More recently, to effectively enhance sequencing coverage across the functional transcript pool, Stewart *et al* developed a universal rRNA-subtraction protocol that was shown to physically remove large amount of rRNA molecules from RNA samples, reducing rRNA transcript abundance by 40-58% (Stewart et al., 2010). The implications of these metatranscriptomic studies are clear: although the sequencing of microbial community transcripts has just begun and is far from comprehensive, it complements the metagenomic approach and has already yielded valuable information on the active components of microbial genomes.

Here we analyze coupled metatranscriptomic and metagenomic data from four bacterioplankton samples taken at Station ALOHA, along the stratified water column characterized by warm, nutrient-depleted surface waters underlain by a steep pycnocline and nutricline (Dore & Karl, 1996; Karl & Lukas, 1996). The goal was to assess in parallel microbial metabolic potential (in DNA) and functional gene expression (in cDNA) along the vertical gradient. In addition to the recent use of these data sets to search and compare putatively novel RNA regulatory elements (small RNAs) highly abundant in these habitats (Shi et al., 2009), the results here demonstrate that coupled metagenomic and metatranscriptomic analyses provide useful perspectives on microbial activity, biogeochemical potential, and regulation in indigenous

microbial populations.

Methods

Sample Collection

Bacterioplankton samples (size fraction 0.22 μm – 1.6 mm) from the photic zone (25m, 75m, 125m) and the mesopelagic zone (500m) were collected from the Hawaii Ocean Time-series (HOT) Station ALOHA site in March 2006, as described previously (Shi et al., 2009). See Supplementary Methods for further details on the seawater collection and RNA/DNA extraction.

Complementary DNA (cDNA) synthesis and sequencing

The synthesis of microbial community cDNA from small amounts of mixed-population microbial RNA was performed as previously described (Frias-Lopez et al., 2008). Briefly, ~100 ng of total RNA was amplified using MessageAmp II (Ambion, Foster City CA) following the manufacturer's instructions and substituting the T7-BpmI-$(dT)_{16}$VN oligo in place of the oligo(dT) supplied with the kit. The SuperScript Double-Stranded cDNA Synthesis Kit (Invitrogen) was used to convert amplified RNA to microgram quantities of cDNA, which was then digested with BmpI to remove poly(A) tails. Purified cDNA was then directly sequenced by pyrosequencing (GS20). See Supplementary Methods for further details.

Bioinformatic analyses

Ribosomal RNA sequences were first identified by comparing the data sets to a combined 5S, 16S, 18S, 23S, and 28S rRNA database derived from available microbial genomes and sequences from the ARB SILVA LSU and SSU databases (www.arb-silva.de). 16S rRNA reads were further selected and subjected to taxonomic classification. Non-rRNA sequences were compared to NCBI-nr, SEED, and GOS protein clusters databases using BLASTX for functional gene analyses as previously described (Frias-Lopez et al., 2008; Shi et al., 2009). Two custom databases (one nucleotide and one amino acid) were constructed from then publicly available 2067 microbial genome sequences, and were used to recruit cDNA and DNA reads. See Supplementary Methods for further details.

Data deposit

The nucleotide sequences are available from the NCBI Sequence Read Archive under

accession numbers SRA007802.3, SRA000263, SRA007804.3 and SRA007806.3 corresponding to cDNA sequences, and SRA007801.5, SRA000262, SRA007803.3 and SRA007805.4 corresponding to DNA sequences, for 25m, 75m, 125m and 500m samples, respectively.

Results and Discussions

Bacterioplankton samples and pyrosequencing data sets

The four sampling depths represent discrete zones in the water column at Station ALOHA (22°45' N, 158°W), which includes the middle of the mixed layer (25m), the base of the mixed layer (75m), the deep chlorophyll maximum (DCM, 125m) at the top of the nutricline, and the upper mesopelagic zone (500m). On cruise HOT179, bacterioplankton samples were collected from each depth for RNA and DNA extraction and sequencing. Since the sampling times for these four sets of seawater samples were different (25m at 22:00 local time, 75m at 03:00, 125m at 06:00, and 500m at 06:00), we expected that the observed gene expression patterns would reflect spatial geochemical gradients (Supplementary Figure S1), as well as temporal differences (discussed below).

A total of ~38 Mbp and ~157 Mbp of sequences were obtained for the four metatranscriptomic and four metagenomic data sets, respectively (Table 1). The number of cDNA reads per GS20 run is roughly a quarter of that of the DNA reads, likely due to incomplete removal of poly(A) tags added during RNA amplification step (Frias-Lopez et al., 2008). Subsequent to the work reported here, significant improvements have been made in the cDNA preparing and sequencing protocols, using the GS-FLX platform (Stewart et al., 2010). Nevertheless, these earlier datasets reported here represent the first set of coupled metagenomic and metatranscriptomic datasets, and provide new information of gene expression in parallel with community structure, gene abundance, and genetic variation.

Taxonomic composition: ribosomal RNA (rRNA) sequence-based analyses

Roughly 0.3% of total DNA reads were designated as rRNA operon sequences (1188, 1117, 954, and 1029 reads for the 25m, 75m, 125m, and 500m samples, respectively), including bacterial, archaeal, and eukaryotic small and large subunit rRNAs, and intergenic spacer

sequences. This sampling frequency was within the expected range based on the rRNA operon size (~5,000 bp), assuming average genome size of ~2 Mbp for marine bacteria and archaea. To assess the taxonomic diversity within the four microbial communities, we classified these 16S rRNA gene sequences (Figure 1, upper panel), using the online Greengenes alignment and classification tools (http://greengenes.lbl.gov/cgi-bin/nph-classify.cgi) (DeSantis et al., 2006), which was reported to yield the highest accuracy for assigning taxonomy to short pyrosequencing reads compared to other methods such as RDP classifier or BLAST (Liu, DeSantis, Andersen & Knight, 2008). These taxonomic assignments were further corroborated (Supplementary Figure S2; Pearson's correlation > 0.95 for all four depths) using a full set of "shotgun" DNA library sequences (average read length 565 bp) from the same source DNA samples (Martinez, Tyson & DeLong, 2010).

Each of the four microbial communities was dominated by two or three major groups (Figure 1, upper panel). Consistiales (predominantly *Pelagibacter*) recruited ~13-35% of the total classified 16S rRNA gene reads from all depths, supporting the high abundance of *Pelagibacter* populations throughout the water column (Eiler, Hayakawa, Church, Karl & Rappé, 2009) and their under-representation in large-insert metagenomic libraries, at least for the populations residing shallower depths (Pham, Konstantinidis, Palden & DeLong, 2008; Temperton et al., 2009). The other major groups included Prochlorales in the photic zone (~17-51%), Cenarchaeales (~22%) and the uncultured delta-proteobacterial group SVA0853 (~9%) at 500m, and Acidimicrobidae (~2-8%) at all depths. This depth distribution was generally consistent with previous cultivation-independent surveys at this site, but variability (likely both biological and methodological) was apparent. For instance, a fosmid library-based survey (DeLong et al., 2006) reported a significant decrease in the relative abundance of *Prochlorococcus* populations at 75m depth, potentially caused by cyanophage infection, as suggested by the large number of cyanophage sequences recovered in the same cellular size fraction. In contrast, in this survey large numbers of phage sequences were not detected, and *Prochlorococcus* relative abundance peaked at 75 m depth, regardless of DNA library type and sequencing method (pyrosequencing, Figure 1; fosmid clone library, Table S1).

Taxonomic composition: Protein-coding sequence-based analyses

Another common approach to assess taxonomic composition from metagenomic data sets

is to infer taxonomic origins from open reading frame (ORF) sequences (Huson et al., 2007). Here, we observed both consistencies as well as some discrepancies when comparing the community composition derived from rRNA gene sequences (discussed above) to those derived from ORF sequences using MEGAN (Huson et al., 2007). As seen in Figure 1 and Supplementary Figure S3, *Pelagibacter* relative abundance decreased from ~13-35% estimated from the 16S rRNA gene sequences, to ~9-23% from the ORF sequences, and the uncultured delta-proteobacterium SVA0853 was completely missed in the latter. In contrast, *Prochlorococcus*-like sequences represented ~39-71% of all annotated ORF sequences, much higher than that estimated from 16S rRNA gene sequences (~17-51%). Higher representation of *Prochlorococcus*-like mRNA transcripts relative to their cell abundance was noted by Poretsky *et al* in metatranscriptomic data sets from day and night samples from the same site, and was attributed to higher transcriptional activities of *Prochlorococcus* cells relative to coexisting heterotrophic microbes (Poretsky et al., 2009). However, it appears that differences in transcriptional activities may not be the explanation, since our DNA data sets showed the same trend of overrepresentation of *Prochlorococcus*-related ORF sequences. Assuming similar genome sizes, a more likely explanation is that the higher representation of *Prochlorococcus*-derived sequences reflects the uneven representation of taxa in current databases. That is, sequence annotation is biased in favor of taxa with more sequenced isolates, such as *Prochlorococcus*, than those with fewer or no sequenced isolates such as *Pelagibacter* and SVA0853-related delta-proteobacteria.

Taxonomic origin of transcripts in the cDNA samples

The simultaneous recovery of rRNA and mRNA transcripts from RNA samples provided a unique opportunity to assess the contribution of each taxon to the community metabolic processes (as judged by transcript abundance). We performed taxonomic analyses with the 16S rRNA as well as protein-coding mRNA transcript sequences exactly as described above for DNA samples (Figure 1, lower panel; Supplementary Figure S3, lower panel). *Prochlorococcus* populations inhabiting DCM layer (125m) displayed highest transcriptional activity, relative to their abundance at that depth. In contrast, *Pelagibacter*, the most numerically abundant heterotrophic bacteria in the open ocean, appeared to be relatively more abundant in cell numbers but less active transcriptionally within DCM layer (also evident in the *Pelagibacter*

genome-wide gene expression analysis below). The DCM layer is characterized by two opposing resource gradients: light supplied from above and nutrients supplied from below, and thus co-existing photoautotrophic and heterotrophic microbes might alternate dominance at different times of a day or in different seasons of a year. Specifically, this apparently lower transcriptional activity of *Pelagibacter* may be influenced by the time of DCM sample collection: ~6AM local time, when photosynthetic microorganisms such as *Prochlorococcus* may be relatively more active.

Finally, for the relatively under-studied mesopelagic zone (500m), two observations are clear. Marine group I crenarchaeota and *Pelagibacter* constitute a major fraction of microbial community both by abundance and metabolic activity. Meanwhile, groups in lower abundance such as *Alteromonadales* and *Sphingomonadales* showed a dramatically higher transcript per gene ratio, suggesting that these groups exhibit higher transcriptional activity than expected based on their DNA abundance.

Global analysis of metabolic potential and functional activities

The majority of the non-rRNA cDNA reads (> 50%), especially those derived from the 500m sample (> 70%), did not share any significant match against NCBI non-redundant (NCBI nr) and the SEED (Meyer et al., 2008) databases (Table 1). Not surprisingly, a significantly higher fraction of cDNA reads shared homology to sequences in the Global Ocean Sampling (GOS) peptide database, the largest marine-specific sequence database available (Yooseph et al., 2007). Furthermore, a large fraction of these cDNA sequences were not present in the coupled DNA libraries at the current sequencing depth (data not shown). These novel sequences likely represented actively expressed ORFs from low abundance microbial groups (alternatively, hyperdynamic genomic regions of well known taxa), or noncoding regions that by definition are not translated into proteins but instead function as RNA molecules (Shi et al., 2009).

For sequences that were annotated as protein coding, we compared gene and transcript abundance in parallel, in order to investigate gene expression in a normalized fashion (see Supplementary Methods). Such normalization accounts for differences in community structure and gene content among samples, allowing detection of metabolic pathways and gene families in lower abundance but with relatively high transcriptional activity (see the example of crenarchaeal-mediated ammonia oxidation at 125m below).

Known metabolic pathways. Several metabolic pathways exhibited high expression levels, as evidenced by a number of SEED subsystems that were found significantly enriched (at the 98% confidence level) in each transcript library, relative to the corresponding DNA library (Figure 2; Table 2). In the surface sample (25m) collected at 22:00 local time, the active expression of oxidative stress-related genes was likely a result of high UV doses during daytime. Aerobic respiration, expected to be enriched relative to photosynthesis at night, was reflected in the expression of cytochrome c oxidases and menaquinone-cytochrome c reductase complexes. The sample collected from DCM layer (125m) at 6:00 AM local time, exhibited high abundance of transcripts associated with carbon fixation and photosynthesis, compared with the other two photic zone samples (despite the relatively lower abundance of photosynthetic genes in the DNA, see Table 2). This is consistent with laboratory observations where *Prochlorococcus* carbon fixation genes were maximally expressed at dawn, and photosynthetic gene expression was elevated upon the appearance of light (Zinser et al., 2009). Highly expressed subsystems in the mesopelagic sample (500m) included peptidoglycan biosynthesis that may be involved in maintenance of cell wall integrity at greater depths, and ammonia assimilation that plays a significant role in energy metabolism for mesopelagic crenarchaeota (Konneke et al., 2005).

Not surprisingly, light-harvesting cellular subsystems were among the most highly expressed in the photic zone. The differentiated clustering of photic zone DNA and cDNA samples observed (Figure 2; Supplementary Figure 5) may be partly attributable to sampling times, given the commonality of diel rhythms among photosynthetic microbes (Zinser et al., 2009). As expected, the metabolic signatures of mesopelagic communities suggested completely different modalities, including energy sources, cellular structures, catabolic and anabolic biochemical pathways.

GOS protein families. The recent global ocean sampling (GOS) expedition (Rusch et al., 2007; Yooseph et al., 2007) has greatly expanded our knowledge of open ocean-derived protein families. Among all protein families identified based on sequence similarity clustering, 3,995 protein clusters consisted of only GOS sequences, 1,700 of which have no detectable homology to previously known protein families (Yooseph et al., 2007). Many of these GOS-only protein clusters of unknown functions were detected in our transcript libraries, some in high abundance (Figure. 3A), underscoring ecologically relevant functions associated with these

novel/hypothetical protein families. Meanwhile, analysis of protein families with known or predicted functions highlighted genes that are highly expressed and therefore likely play active roles in maintaining ecosystem functions at each habitat (Figure 3B).

Nitrogen metabolism protein families. A suite of nitrogen metabolism genes (ammonium transporter, *amt*; dissimilatory nitrite reductase, *nirK*; urea transporter, *urt*; ammonia monooxygenase subunits, *amoABC*) was among the most highly expressed of GOS protein families detected (Figure 3B). An essential macronutrient, nitrogen availability and turnover limits biological production in many open ocean regions, including NPSG (Van Mooy & Devol, 2008). Ammonia/ammonium is a key reduced nitrogen compound that can either be incorporated into carbon skeleton via the glutamine synthetase (GS; *glnA*)/glutamate synthase (GOGAT; *glsF*) cycle, or can serve as energy source fueling autotrophic metabolism (Konneke et al., 2005). Thus, the transport of ammonia/ammonium is vital to planktonic microbes living in the nutrient deplete surface waters and energy constrained deep waters in an open ocean setting. Urea is another potentially important nitrogen source in the ocean, and is utilized by marine cyanobacteria (Moore, Post, Rocap & Chisholm, 2002). The more oxidized forms of nitrogen, nitrite and nitrate require more metabolic energy to utilize but can serve as alternative nitrogen sources because of their much higher concentrations in deep euphotic zone and mesopelagic zone below the nitracline.

To assess the prevalent nitrogen utilizing pathways in the genomes of the most abundant planktonic microbial populations, we compared the observed frequency (normalized to gene length and data set size) of several essential nitrogen metabolism genes with that of the 16S rRNA gene of *Prochlorococcus* and marine group I crenarchaeota. The observed frequency of *Prochlorococcus*-related *amt*, *glnA*, *urt*, urease genes is equivalent to that of *Prochlorococcus* 16S rRNA gene (Supplementary Figure S4A, left panel), suggesting that ammonium and urea assimilation is preserved in naturally occurring *Prochlorococcus* populations. In contrast, the assimilatory nitrite reductase gene (*nirA*) was present in only a small fraction of *Prochlorococcus* cells (c.a., 7%, 8% and 15% at 25m, 75m, and 125m, respectively), consistent with expectation based on genomic and physiological studies of *Prochlorococcus* isolates (Moore et al., 2002; Rocap et al., 2003). Furthermore, the transcripts of these nitrogen metabolism genes (except *nirA*) were also detected in our metatranscriptomic data sets

(Supplementary Figure S4A, right panel), suggesting active deployment of these nitrogen metabolism pathways by *Prochlorococcus* cells *in situ*. The *amt* gene was the most actively transcribed, likely an adaptive mechanism to efficiently scavenge low-concentration ammonium as the most preferred nitrogen source. The dramatic decrease in *amt* gene expression at 125m however, was not expected. It is possible that the apparently higher primary production at 125m (DCM) has caused accumulation of ammonium via active nutrient regeneration processes. In fact, ammonium maxima near the DCM layer are common in stratified oligotrophic waters (Brzezinski, 1988). As a result, the presumably elevated ammonium concentration may result in down-regulation of the *amt* gene expression, as observed in many cyanobacteria isolates.

Marine group I crenarchaeota exist in high abundance in mesopelagic zone, where distinct forms and concentrations of nitrogen species (e.g., nitrate, nitrite, urea) are present. *Nitrosopumilus maritimus*, an isolate of related crenarchaea from marine aquarium, has been shown definitively to grow chemolithoautotrophically on ammonia (Konneke et al., 2005). Further genomic analyses of marine group I crenarchaeota have provided insights into the metabolism of other forms of nitrogen compounds (Hallam et al., 2006; Walker et al., 2010). Here, our data showed that *amt*, *amoABC*, and *glnA* genes were prevalent and expressed in planktonic crenarchaeal populations, whereas urea utilization genes, while present and expressed, appeared in lower abundance (Supplementary Figure S4B, left panel). Clearly, despite the apparent lack of such genes in the *N. maritimus* genome (Walker et al., 2010), a fraction of planktonic crenarchaeal populations encode genes for utilizing urea as nutrient or energy source. The normalized expression levels of crenarchaea-related *amt* and *amoABC* genes (especially *amoC* gene) was among the highest in our data sets (orders of magnitude higher than most other protein-coding genes) (Figure 3B). Interestingly, the anomalously high *amoC* gene expression appeared to be universal, as also observed in bacterial nitrifiers (Berube, Samudrala & Stahl, 2007), for as-yet unknown reasons. Consistent with a quantitative PCR-based study (Church et al., 2010), the *amoABC* transcripts were detected in high abundance at 125m depth despite the small planktonic crenarchaeal population size (Supplementary Figure S4B, right panel). Together with previous report of remarkably high substrate affinity and kinetics of crenarcheal *amo* genes (Martens-Habbena, Berube, Urakawa, de la Torre & Stahl, 2009), these data further support a role for marine crenarchaea in nitrification in the ocean via active ammonia oxidation.

Nitrite, an end product of archaeal ammonia oxidation, could exert toxic effects to cells if accumulated, and an upper primary nitrite maximum (UPNM) is often observed near DCM layer (125m in this study) in the open ocean (Dore & Karl, 1996). Consistent with the hypothesis that dissimilatory nitrite reductase (*nirK*) in ammonia-oxidizing microbes is involved in nitrite detoxification (Casciotti & Ward, 2001; Hallam et al., 2006), *nirK* was found highly expressed at 125m (Supplementary Figure S4B, right panel). Finally, nitrate reductase genes (*narH* and *narG*) and transcripts were frequently detected in the 500m data sets, and appeared to be most similar to homologs found in Candidatus *Kuenenia stuttgartiensis* (data not shown), suggesting that planktonic crenarahaea may not participate in the first step of nitrate respiration.

Photoheterotrophy. We detected in the photic-zone active expression of genes involved in photoheterotrophy, including those encoding proteorhodopsins. Proteorhodopsin (PR) is a photoprotein that functions as light-driven proton pump, generating biochemical energy via proton motive force (Béjà et al., 2000). PR photosystems have been detected in a large percentage (up to 80%) of ocean surface-dwelling bacteria and archaea (DeLong & Béjà, 2010), and were suggested to be horizontally transferred among phylogenetically divergent microbial taxa (Frigaard, Martinez, Mincer & DeLong, 2006; McCarren & DeLong, 2007). Laboratory-based experiments have suggested that PR photosystem increases cellular fitness to bacterial cells under adverse growth conditions (González et al., 2008; Gómez-Consarnau et al., 2010; Gómez-Consarnau et al., 2007).

Our depth profile data allow us to directly assess the *in situ* abundance and taxonomic origins of PR gene and transcripts. Abundance of PR transcripts decreased dramatically from euphotic zone to 500m (in which only 4 cDNA reads shared homology to known PR genes) (Supplementary Figure S5A). While PR DNA and cDNA reads appeared to be originated from a diverse range of taxa, the majority shared homology to known PR genes from SAR11-like organisms (Supplementary Figure S5B). Notably, PR genes were found most highly expressed in the 75m sample (collected at 22:00), followed by the 25m and 125m samples (collected at 3:00 and 6:00, respectively) (Supplementary Figure S5A; also see the *Pelagibacter* genome-wide gene expression analysis below), suggesting PR genes may be constitutively expressed in the photic zone independent of light conditions. Laboratory studies of PR-containing isolates as well as a recently reported microcosm experiment have reported inconsistent observations, some

suggesting constitutive PR expression (Giovannoni et al., 2005a; Riedel et al., 2010), while others suggesting light-regulation of PR expression (Gómez-Consarnau et al., 2007; Lami, Cottrell, Campbell & Kirchman, 2009). Higher-resolution metatranscriptomic studies are necessary to provide further insight into light effects on PR gene expression in different taxa, and in different oceanographic provinces.

Evidence for another form of phototrophy mediated by aerobic anoxygenic phototrophic (AAP) bacteria was also observed. Recent studies suggest that AAPs constitute a considerable fraction of marine planktonic community, and may contribute significantly to the carbon cycle in the ocean via facultative photoheterotrophy (Béjà et al., 2002; Kolber et al., 2001). Living in an oligotrophic environment, oceanic AAPs likely are capable of efficiently controlling the expression of their photosynthetic apparatus, supplementing heterotrophic metabolism with light-dependent energy harvest. In this depth profile, AAPs were most abundant in 25m and 75m samples based on observed gene frequencies of bacteriochlorophyll biosynthesis genes (*bchXYZ*), light-harvesting complex I genes (*pufAB*) and the reaction center genes (*pufLM*). The majority of these photosynthetic genes were closely related to *Roseobacter*-like AAP sequences, particularly a BAC clone insert retrieved from the Red Sea (eBACred25D05; accession number: AY671989) (Oz et al., 2005). GOS protein clusters associated with these AAP genes were found highly expressed in the 75m sample (Figure 3B), and most of this AAP gene expression originated from the *puf* operon (Supplementary Figure S6). Collectively, the data indicate photosynthetically active population of AAPs, at 75m in particular.

Reference genome-centric analyses

We used a total of 2067 genomic references (including finished and draft genomes), to recruit DNA and cDNA reads at high stringency, based on BLASTN comparison (see Supplementary Methods). About 29%, 40%, 15% and 7% of total DNA reads, and 30%, 24%, 26%, and 18% of total cDNA reads were recruited to the reference genomic data for 25m, 75m, 125m, and 500m sample, respectively. Notably, the percentage of recruited cDNA reads for each sample was significantly higher than that of cDNA reads that could be assigned to NCBI-nr protein database (Table 1), a result of cDNA recruitment to expressed noncoding regions on the genomes. For instance, about 1539 reads in the 25m sample were recruited to an intergenic region of *Prochlorococcus* strain MIT 9215 genome, corresponding to the Group_2 small RNA

previously reported by Shi *et al* (Shi et al., 2009).

The relative representation of genomes/genome fragments is shown in a three-way comparison plot, to illustrate the similarities and differences of communities dwelling in specific habitats (Figure 4). For this analysis, the 75m and 125m samples were pooled together, since they share similar profile at both DNA and cDNA levels (Figure 2). All genomes recruiting > 50 DNA reads are also listed in Supplementary Table S2. Here, general separation of photic zone populations with mesopelagic populations was observed, with a few exceptions that were found more evenly distributed along the depth, including the ubiquitous *Pelagibacter*, and the alphaproteobacterium *Erythrobacter* sp. SD-21, a Mn(II) oxidizing bacterium that has been isolated from many diverse marine environments including surface and deep oceans (Francis, Co & Tebo, 2001).

Such genome recruitment analysis provides direct measurement of vertical distribution of ecologically coherent populations (represented by reference genomes) in nature, such as high-light (HL) and low-light (LL) adapted *Prochlorococcus* "ecotypes" (Moore & Chisholm, 1999). Notably, despite an expected significant increase of low-light (LL) adapted *Prochlorococcus* populations (mostly eNATL2A) at 125m, where light intensity dramatically decreased compared to shallower depths, > 80% of the *Prochlorococcus*-like reads at 125m were most similar to sequences of high-light (HL) adapted isolates (mostly eMIT9312) (Supplementary Table S2). While possibly a result of physical homogenization of the water column due to deep mixing in the winter (Malmstrom et al., 2010), these HL-like *Prochlorococcus* cells displayed elevated transcriptional activity at 125m (Supplementary Table S2), suggesting they were unlikely sinking dead cells. Zinser and colleagues (Zinser et al., 2006) showed that in deeper waters (below 75 m) at the western North Atlantic site, a significant fraction of *Prochlorococcus* population cannot be detected by qPCR probes designed to capture currently known ecotypes, suggesting significant deep populations of *Prochlorococcus* yet to be identified and characterized. Results here suggest the presence of a HL-like *Prochlorococcus* population that may be well adapted to the lower euphotic zone, under low light conditions.

Population transcriptomic analysis of *Pelagibacter*. As the most abundant heterotrophic bacterial group throughout the ocean water column, *Pelagibacter* (member of the alphaproteobacteria SAR11 clade) provides a useful model example for how culture-based and

metagenomic/metatranscriptomic data can be integrated to study the ecophysiology of wild populations. Subsets of DNA and cDNA reads from all 4 depths were mapped onto the reference genome of the open ocean *Pelagibacter* isolate HTCC7211 (see Supplementary Methods). The expression level of annotated protein coding genes provided clues on the prevailing metabolic activities of *Pelagibacter* populations at each depth (Figure 5; Supplementary Table S3). Overall, the expression profile of protein coding genes confirmed the observation based on the rRNA profile (Figure 1), that *Pelagibacter* cells at 125m were less transcriptionally active at the time of sampling, compared to their counterparts at 25m and 75m. Indeed, ribosomal proteins were among the most highly expressed genes in 25m and 75m samples, and most ORFs showed lower expression levels in the 125m sample.

Nutrient-uptake genes of *Pelagibacter*, particularly those encoding periplasmic solute binding proteins of ATP-binding cassette (ABC) families, represented the most abundant class of transcripts (Figure 5). The disproportionally high abundance of transporter genes in *Pelagibacter* genomes is believed to contribute to their capability of efficiently utilizing a broad variety of substrates (Giovannoni et al., 2005b). Here we observed high transcriptional levels of solute-binding proteins families 1, 3, and 7 (Figure 5), which involve in the uptake of sugars, polar amino acids, and organic polyanions, respectively (Tam & Saier, 1993). Polyamines (e.g., spermidine/putrescine), trace elements (e.g., selenium), and possible osmolytes (e.g., glycine betaine) also appeared to be actively transported. In addition, a few transporter families other than the ABC superfamily were also expressed, including Na+/solute symporter (Ssf family) and tripartite ATP-independent periplasmic (TRAP) dicarboxylate transporter genes for the uptake of mannitol and/or C4-dicarboxylates, which relies on proton motive force rather than ATP hydrolysis. Notably, different expression levels among the four depths were discernible for these transporter genes, potentially a result of substrate availability and preference for *Pelagibacter* populations residing different depths.

Sowell and colleagues have observed in *Pelagibacter* metaproteomes collected from the Sargasso Sea surface water a dominant signal of periplasmic transport proteins for substrates such as phosphate, amino acids, phosphonate and spermidine/putrescine (Sowell et al., 2008). The overall consistent observation that nutrient-uptake transporters were most highly expressed both at transcriptional level (this study) and translational level (Sowell et al., 2008), corroborates

the oligotrophic nature of both oceanic sites. However, significant differences in peptide versus transcript expression levels were also apparent among certain categories of transporters. For example, we did not detect gene expression for phosphate and phosphonate transporter genes (*pstS* and *phnD*) related to *Pelagibacter* in our data sets. In fact, no *phnD*-related sequences were detected in the DNA reads recruited to the *Pelagibacter* HTCC7211 genome, suggesting *phnD* gene is absent in most *Pelagibacter* cells at Station ALOHA. This observation contrasts sharply with the that of Sowell *et al*, reflecting the significant biogeochemical difference between the two oceanic sites (e.g., phosphate concentrations at BATS are much lower than that at Station ALOHA (Wu, Sunda, Boyle & Karl, 2000)). The effect of geography-dependent phosphorus limitation appears to be reflected in the gene content of native *Prochlorococcus* cells (Martiny, Huang & Li, 2009), as well as other picoplankton populations (Martinez et al., 2010).

HTCC7211-specific genes. It has been well established that genomic plasticity of microbes, reflected by variations in gene content of closely related strains, may facilitate microbial adaptation to their natural habitats (Coleman et al., 2006; Cuadros-Orellana et al., 2007). We compared the genome sequences of two *Pelagibacter* coastal isolates (strains HTCC1062 and HTCC1002) and the open ocean isolate (HTCC7211, used as reference genome in the genome-centric analysis above), and asked which HTCC7211-specific genes might be highly expressed and thus functionally important in the open ocean environment.

There are 296 HTCC7211-specific genes (see Supplementary Methods), 154 detected in at least one of our metatranscriptomic data sets (Supplementary Figure S7). Two ORFs encoding ABC-type periplasmic solute binding proteins appeared to be specific to open ocean-dwelling *Pelagibacter*, and were highly expressed. One ORF encodes a selenium-binding protein, which may contribute to the synthesis of selenoproteins (Zhang & Gladyshev, 2008). The other ORF encodes an extracellular solute-binding protein family 1, which is associated with the uptake of malto-oligosaccharides, multiple sugars, alpha-glycerol phosphate, and iron (Tam & Saier, 1993). In addition, the C4-dicarboxylate transport (Dct) system, which relies on highly specific and affine extracytoplasmic solute binding receptors, appeared to be important in oceanic *Pelagibacter* populations. Not only were four *dct* operons present in the strain HTCC7211 (as opposed to apparently only one copy in coastal strains HTCC1062 and HTCC1002), but the three HTCC7211-specific *dctP* paralogues (encoding a periplasmic C4-dicarboxylate-binding

protein) were also expressed (Supplementary Figure S7). Dct transporters are secondary carriers that use an electrochemical H^+ gradient as the driving force for transport rather than ATP hydrolysis, and allow the uptake of mannitol and/or C4-dicarboxylates like succinate, fumarate, and malate, pointing to such organic compounds as important carbon and energy source for oceanic *Pelagibacter*.

Tables and Figures

Table 1. Summary of 4 metagenomic data sets and 4 metatranscriptomic data sets.

HOT 179	Depth	# of total reads	Ave. read length (bp)	# of rRNA reads	% of rRNA in total reads	# of non rRNA reads	hits to protein db (% of non rRNA)			
							COG	SEED	NCBI-nr	GOS protein family
cDNA	25 m	74638	99	33878	45.4	40760	7.5	11.2	17.1	45.3
	75 m	106936	99	62096	58.1	44840	6.0	9.9	15.3	49.4
	125 m	97915	97	45809	46.8	52106	6.2	10.4	16.1	46.2
	500 m	109249	97	40537	37.1	68712	3.8	4.4	10.1	26.3
DNA	25 m	359665	109	1188	0.3	358477	19.1	26.7	42.0	63.5
	75 m	388652	110	1117	0.3	387535	22.4	33.2	51.3	71.9
	125 m	322751	109	954	0.3	321797	18.1	23.4	36.3	60.9
	500 m	371071	107	1029	0.3	370042	17.3	18.3	30.5	49.0

Table 2. SEED subsystems that are significantly enriched in cDNA data sets relative to DNA data sets (0.98 confidence level, based on the method described in Rodriguez-Brito *et al.*, 2006).

Depth	Subsystem*	Representation in cDNA	Representation in DNA
25m	Ammonia_assimilation	1.52%	0.25%
	Photosystem_I	1.72%	0.58%
	Proteorhodopsin	1.00%	0.03%
	Ribosome_LSU_bacterial	3.04%	1.23%
	Ribosome_SSU_bacterial	2.58%	0.79%
	Universal_GTPases (mostly elongation factors)	2.36%	1.31%
	RNA_polymerase_bacterial	2.46%	1.25%
	Transcription_initiation,_bacterial_sigma_factors	0.80%	0.21%
	Terminal_cytochrome_C_oxidases	1.60%	0.38%
	Ubiquinone_Menaquinone-cytochrome_c_reductase_complexes	0.58%	0.11%
	Oxidative_stress	0.90%	0.28%
75m	Ammonia_assimilation	1.09%	0.26%
	Photosystem_I	2.38%	0.66%
	Photosystem_II	2.31%	0.81%
	Proteorhodopsin	0.80%	0.03%
	Ribosome_LSU_bacterial	2.90%	1.20%
	Ribosome_SSU_bacterial	1.97%	0.79%
125m	CO2_uptake,_carboxysome	1.20%	0.49%
	Peptidoglycan_Biosynthesis	2.28%	1.24%
	Chlorophyll_Biosynthesis	2.34%	0.87%
	Photosystem_I	5.24%	0.37%
	Photosystem_II	5.21%	0.46%
	Proteorhodopsin	1.34%	0.04%
	Ribosome_LSU_bacterial	3.92%	1.32%
	Ribosome_SSU_bacterial	2.69%	0.77%
	Universal_GTPases (mostly elongation factors)	3.05%	1.37%
	F0F1-type_ATP_synthase	2.14%	0.92%
	Cytochrome_B6-F_complex	0.86%	0.16%
	Transport_of_Iron	1.78%	0.40%
500m	Peptidoglycan_Biosynthesis	4.63%	1.12%
	Ammonia_assimilation	3.43%	0.12%
	Ribosome_SSU_bacterial	1.41%	0.67%
	Terminal_cytochrome_C_oxidases	1.55%	0.51%

* Subsystems listed are significantly enriched in cDNA samples at the 0.98 confidence level

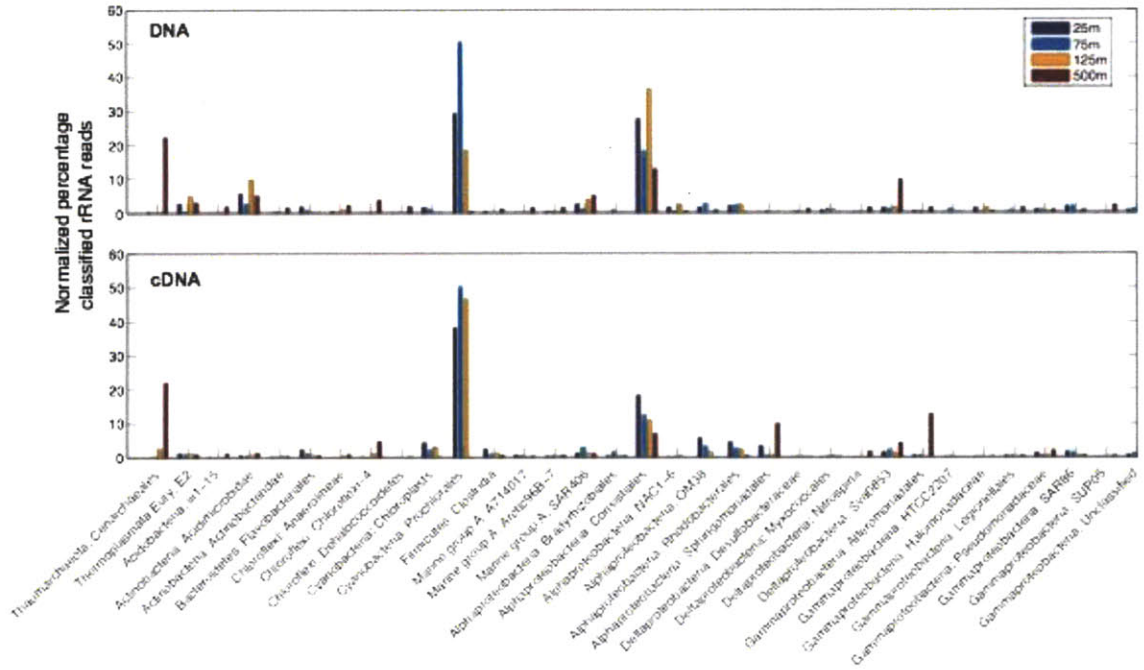


Figure 1. Taxonomic classification based on 16S rRNA-bearing reads in DNA and cDNA data sets. Taxonomic assignments were binned at the Order level, using the Hugenholtz taxonomy of Greengenes (see Supplementary Methods). 16S rRNA sequences that could not be classified were excluded from the analysis. Y-axis scale represents the percentage of the total classified 16S rRNA reads. Only taxa that represented $\geq 1\%$ of all classified reads are displayed. Also note here that, since no replicate data were available for each sample, error bars were absent and thus no statistical inference could be made from the figure.

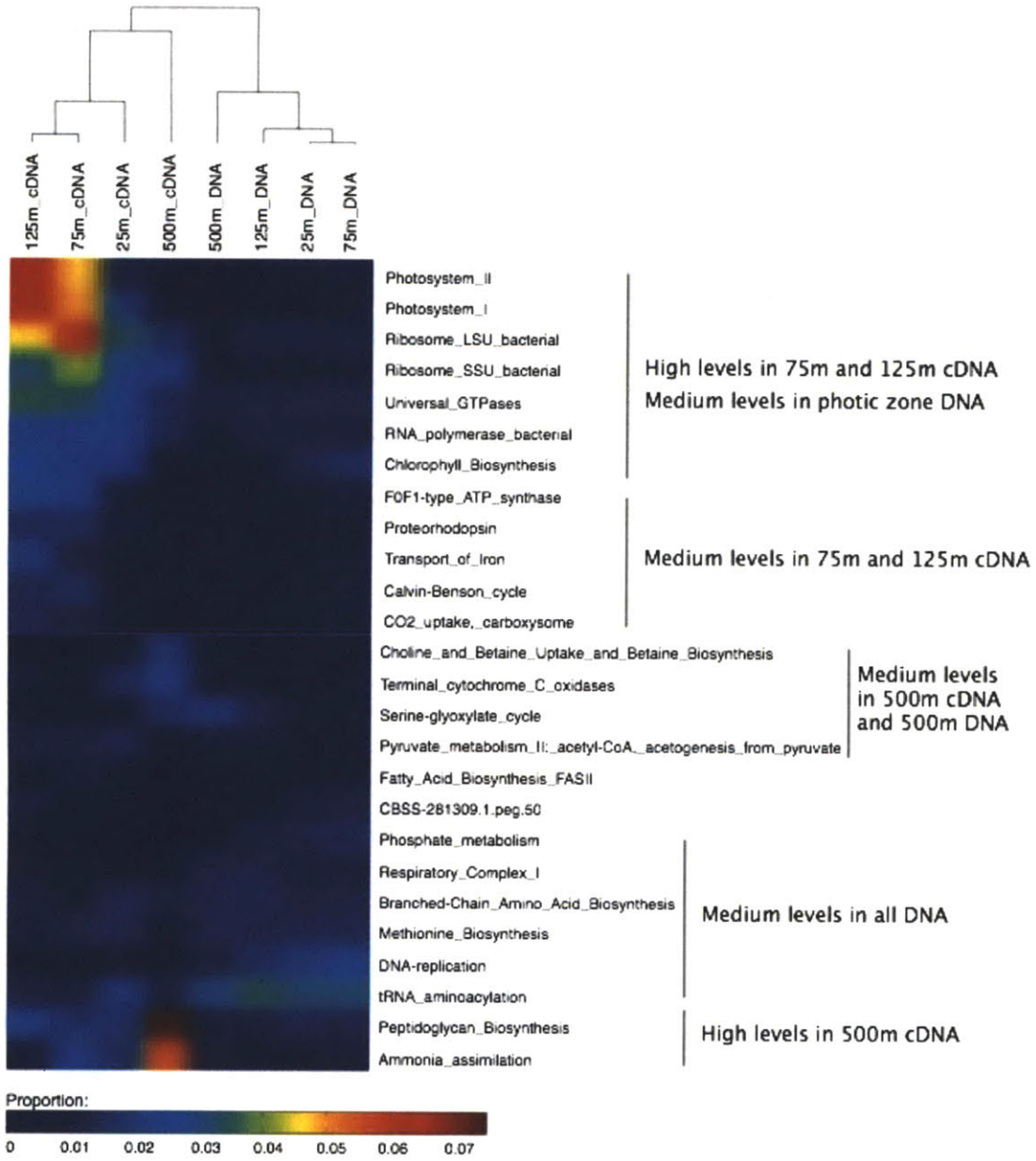


Figure 2. Clustering of all cDNA and DNA data sets based on relative abundance of SEED subsystems. Only the most abundant subsystems that together recruited 95% of all reads are displayed. Hierarchical clustering of 4 DNA and 4 cDNA samples were performed with euclidean distance and single linkage method using MATLAB. Color scale represents the proportion of reads assigned to SEED categories relative to the total library size in each sample. Blue to red color indicates low to high representation of SEED categories.

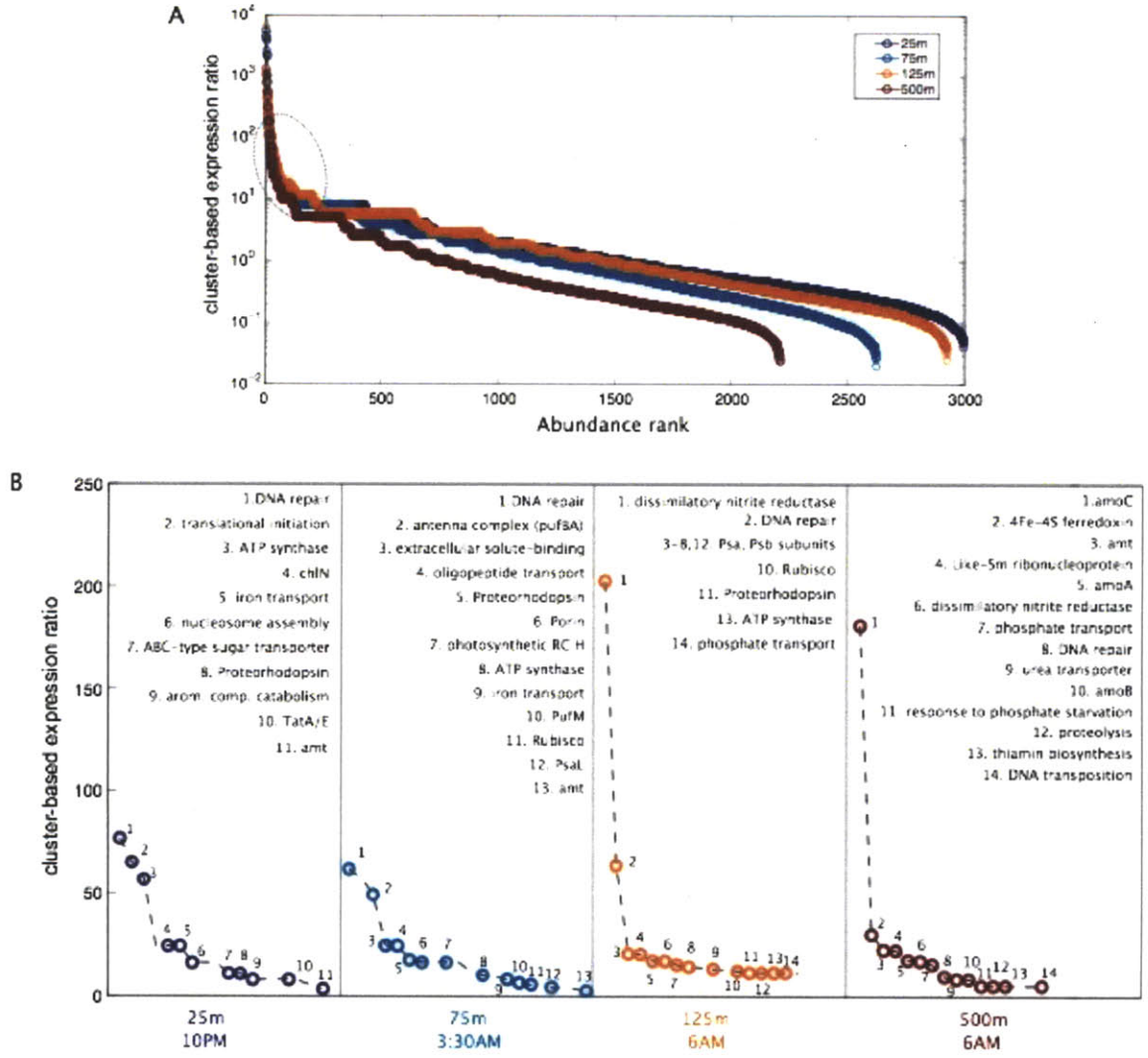


Figure 3. Community-level gene expression profiles based on the GOS protein family database. Cluster-based expression ratio was defined as representation of each GOS cluster in the cDNA library normalized by its representation in the DNA library. GOS clusters that recruited only cDNA reads were arbitrarily set a value of 1 copy of DNA read, to avoid a denominator of 0. (A) GOS clusters were ranked by their cluster-based expression ratios for four depths; (B) The most highly expressed GOS clusters with known or predicted functions were highlighted for each depth.

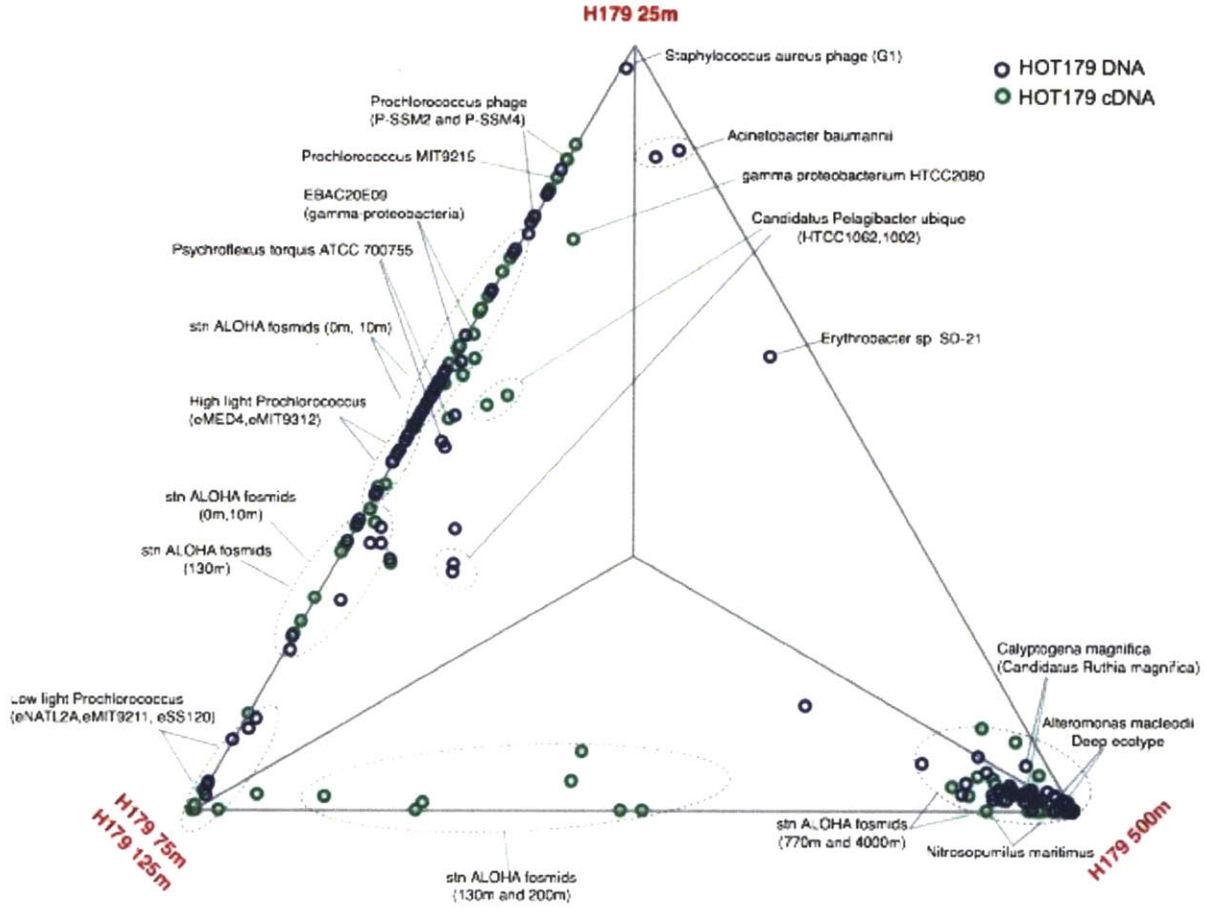


Figure 4. Three-way comparison of representation of genomes and genome fragments (fully sequenced fosmids) in DNA and cDNA data sets. The 75m and 125m data sets were combined since they were the most similar. Each dot represents a genome (fragment), and its proximity to a vertex reflects the enrichment of the corresponding genome (fragment) in the respective sample. Only genomes recruited > 0.1% of total reads are displayed. Station ALOHA fosmids represent fosmid sequences that were reported by DeLong et al (DeLong et al 2006). See Supplementary Methods for detail.

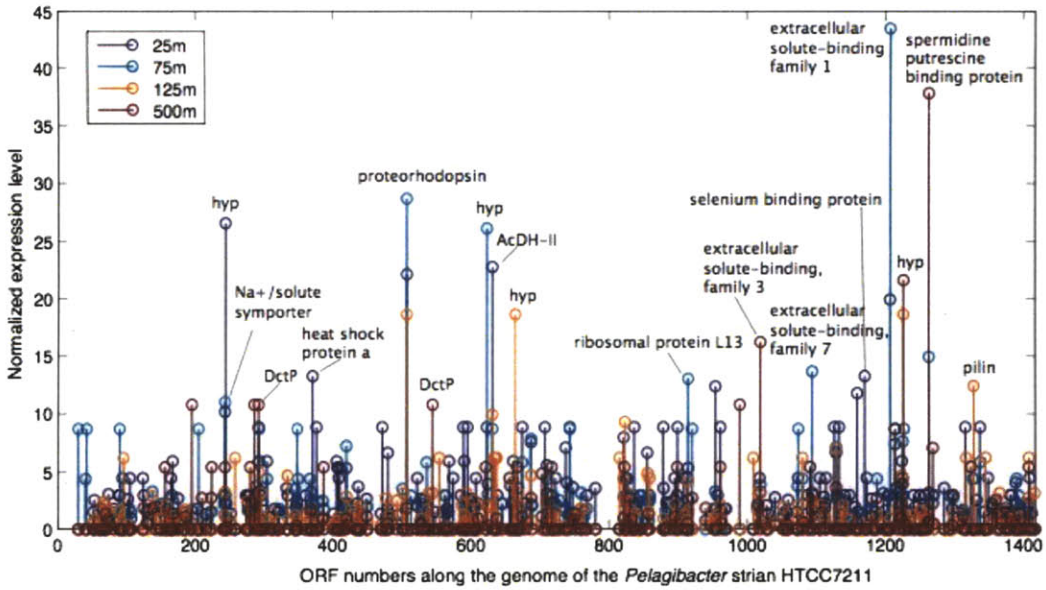


Figure 5. Genome-wide expression profiles of *Pelagibacter*-related populations, in all four depths. X-axis shows the arbitrary numbering of ORFs along the genome of *Pelagibacter* strain HTCC7211. Y-axis scale represents normalized cDNA to DNA ratio (normalized expression level; see Supplementary Methods) for each ORF. Each colored circle in the stem plot represents a given ORF at a given depth.

Acknowledgements and author contributions:

We thank the captain and crew of the R/V Kilo Moana for facilitating sample collection. Thanks also to Stephan Schuster for collaboration on pyrosequencing. We are grateful to the J. Craig Venter Institute, and the Gordon and Betty Moore Foundation for the microbial genome sequences. This work was supported by the Gordon and Betty Moore Foundation, National Science Foundation Microbial Observatory Award MCB-0348001, the Department of Energy Genomics GTL Program, and the Department of Energy Microbial Genomics Program, and an NSF Science and Technology award, C-MORE.

Supplementary Information for Chapter 3

Supplementary Methods
Supplementary Tables S1-S3
Supplementary Figures S1-S8

Supplementary Methods

Sample Collection and DNA/RNA extraction

Bacterioplankton samples (size fraction 0.22 μm – 1.6 mm) from the photic zone (25m, 75m, 125m) and the mesopelagic zone (500m) were collected from the Hawaii Ocean Time-series (HOT) Station ALOHA site in March 2006, as described previously (Shi et al., 2009). Briefly, four replicate 1-liter seawater samples were prefiltered through 1.6-mm GF/A filters (Whatman, Maidstone, U.K.) and then filtered onto 0.22-μm Durapore filters (25mm diameter, Millipore, Bedford, MA) using a four-head peristaltic pump system. Each Durapore filter was immediately transferred to screw-cap tubes containing 1 ml of RNAlater (Ambion Inc., Austin, TX), and frozen at -80°C aboard the R/V Kilo Moana. Samples were transported frozen to the laboratory in a dry shipper and stored at -80°C until RNA extraction. Total sampling time, from arrival on deck to fixation in RNAlater was less than 20 minutes.

Replicate filters were pooled for RNA extractions, which were performed as previously described (Shi et al., 2009), using the *mir*Vana™ RNA isolation kit (Ambion, Austin, TX). Samples were thawed on ice, and the 1 ml RNAlater was loaded onto two Microcon YM-50 columns (Millipore, Bedford, MA) to concentrate and desalt each sample. The resulting 50 l of RNAlater was added back to the sample tubes, and total RNA extraction was performed following the *mir*Vana™ manual. Genomic DNA was removed using a Turbo DNA-free™ kit (Ambion, Austin, TX). Finally, extracted RNA (DNase-treated) from four replicate filters were combined, purified, and concentrated by using the MinElute PCR Purification Kit (Qiagen, Valencia, CA).

Bacterioplankton sampling for DNA extraction and DNA extraction was performed as

previously described (Frias-Lopez et al., 2008).

RNA amplification and cDNA synthesis

Roughly 100 ng of total RNA was amplified using the MessageAmp II-Bacteria kit (Ambion) as described previously (Frias-Lopez et al., 2008) (Shi et al., 2009). Briefly, total RNA was polyadenylated using Escherichia coli poly(A) polymerase. Polyadenylated RNA was converted to double-stranded cDNA via reverse transcription primed with an oligo(dT) primer containing a promoter sequence for T7 RNA polymerase (underlined) and a recognition site for the restriction enzyme BpmI (T7-BpmI-$(dT)_{16}$VN; nt sequence: GCCAGTGAATTGTAATACGACTCACTATAGGGGCGACTGGAGTTTTTTTTTTTTT TTTVN). cDNA was then transcribed in vitro at 37 °C for 6 hours, yielding large quantities (~100 ug) of antisense RNA. An aliquot of antisense RNA (~5 ug aliquot) was polyadenylated again and converted to double-stranded cDNA using first the SuperScript III First-Strand Synthesis System (Invitrogen, Carlsbad, CA, USA) with priming via oligo(dT) for first-strand synthesis, and then the SuperScript Double-Stranded cDNA synthesis kit (Invitrogen) for second-strand synthesis. cDNA was then purified with the QIAquick PCR purification kit (Qiagen), digested with BpmI for 2-3 hours at 37 °C to remove poly(A) tails, and purified again with the QIAquick PCR purification kit. Purified cDNA was used for the generation of single-stranded DNA libraries and emulsion PCR according to established protocols (454 Life Sciences, Roche). Clonally amplified library fragments were then sequenced on a Genome Sequencer GS20 System (Roche).

Bioinformatics analyses

Taxonomic classification of 16S rRNA sequences. Ribosomal RNA sequences were first identified by comparing the data sets to a combined 5S, 16S, 18S, 23S, and 28S rRNA database derived from available microbial genomes and sequences from the ARB SILVA LSU and SSU databases (www.arb-silva.de). 16S rRNA sequences were then selected by BLASTing (Altschul et al., 1990) against SILVA SSU databases (bits score $\geq$ 50, alignment length $\geq$ 80% of the read length, and alignment length $\geq$ 100bp), and classified using the online Greengenes classifier tools

(http://greengenes.lbl.gov/cgi-bin/nph-classify.cgi), using the Hugenholtz taxonomy. The parameters used for classifying 16S rRNA were a minimum alignment length of 100bp, and a minimum sequence identity of 75%. For the shotgun sequences, 16S rRNA reads were chosen based on the cutoff of a bits score ≥ 50 and an alignment length ≥ 280bp, and the parameters used for classifying 16S rRNA were a minimum alignment length of 280bp, and a minimum sequence identity of 75%.

Taxonomic classification of protein-coding sequences. Protein-coding sequences were identified by blasting against the NCBI non-redundant (NCBI-nr) protein database. The BLASTx output was parsed to analyze the taxonomic breakdown using MEGAN (Huson et al., 2007), with bit scores > 40 within 10% of the top scoring hits.

Functional analyses using the SEED database and GOS protein cluster database. Non rRNA reads were assigned to SEED subsystems and GOS protein clusters based on top BLASTx hits with bits score ≥ 40. A bootstrapping method (Rodriguez-Brito, Rohwer & Edwards, 2006), which takes care of the size difference among subsystems and looks for statistically significant differences metagenomes, was applied to identify subsystems that were enriched in the cDNA libraries relative to the corresponding DNA libraries. GOS protein cluster-based analysis was perform as previously described (Frias-Lopez et al., 2008). Briefly, cluster-based expression ratios were calculated as the number of reads found for each protein cluster in the cDNA library relative to that found in the DNA library, which was further normalized for the difference in DNA and cDNA library size. Functional annotations for GOS protein clusters, when available, were available from a study by Yooseph *et al* (Yooseph et al., 2007). The cluster-based expression ratios were ranked from highest to lowest (Figure 3) to look at clusters being expressed at elevated levels.

Reference genome-centric analysis. Two custom databases (one nucleotide database and one amino acid database) were constructed from 2067 publicly available microbial genome sequences and annotations (fully sequenced and draft genomes as of January 2009). Non-rRNA cDNA and DNA reads from all four depths were compared against the custom nucleotide database, and reads with top hit bits score ≥ 40 were assigned to the corresponding genome. In

order to compensate for likely uneven phylogenetic representation in the databases, we allowed any read to map to several reference read with the same alignment score. Recruitment of protein-coding cDNA and DNA reads onto reference genomes were performed by assigning reads to top amino acid sequences with bits score ≥ 40. For each ORF, recruited cDNA abundance was divided by the recruited DNA abundance, to give an indication of per-copy cDNA level. If there were cDNA hits but no DNA hits for a given ORF, the number of DNA hits was considered as 1.

To examine the expression of *Pelagibacter* strain HTCC7211-specific ORFs, **p**utative *Pelagibacter* reads were first identified as reads with top BLASTx hit (against NCBI-nr) to *Pelagibacter* and with a bit score >40. Each of these putative *Pelagibacter* reads then was searched against a custom database of *Pelagibacter* ORFs derived from 3 fully sequenced *Pelagibacter* strains (HTCC1062, HTCC1002, HTCC7211) using BLASTx, and assigned to the best hit ORF. The HTCC7211-specific ORFs were identified as ORFs with no best reciprocal hit, based on the cutoff of a minimum sequence identity of 30%, and a minimum alignment length fraction of 75%, in the genomes of HTCC1062 or HTCC1002.

Supplementary Tables and Figures

Table S1. Comparison of Prochlorales representation in HF (DeLong et al, *Science*, 2006) and HOT 179 fosmid clone libraries.

Fosmid library	Sampling depth	# reads assigned to a taxon*	# (%) reads assigned to Prochlorales
HF	10 m	5165	341 (6.6%)
	75 m	5953	124 (2.1%)
	130 m	4530	169 (3.7%)
	500 m	6777	6 (0.09%)
HOT179	25 m	8196	820 (10%)
	75 m	10120	1502 (14.8%)
	125 m	15375	1300 (8.5%)
	500 m	16544	22 (0.13%)

* Taxon breakdown was performed with MEGAN (Huson *et al.* 2007), using the following LCA parameters: min support = 1, min score = 70, top percent = 0.

Table S2. Recruitment of cDNA and DNA reads to abundant reference genomes.

Reference genomes	# of DNA reads assigned to a reference genome				# of cDNA reads assigned to a reference genome			
	25m	75m	125m	500m	25m	75m	125m	500m
Prochlorococcus marinus AS9601	28682	43034	10311	23	1656	1900	1926	4
Prochlorococcus marinus MIT 9301	24272	37042	8733	19	1683	2081	1887	7
Prochlorococcus marinus MIT 9312	14405	22578	5805	12	926	1125	1043	2
Prochlorococcus marinus MIT 9215	14354	21886	5193	21	5039	1902	2225	18
Prochlorococcus marinus MED4	1277	2737	644	5	197	269	163	0
Candidatus Pelagibacter ubique HTCC1062	1137	1241	1642	612	238	204	291	84
Candidatus Pelagibacter ubique B HTCC1002	1102	1242	1616	628	232	196	262	102
Psychroflexus torquis ATCC 700755 ATCC700755	1383	1287	1436	181	170	195	187	30
Prochlorococcus marinus NATL1A	126	847	2571	2	13	43	569	0
Prochlorococcus marinus NATL2A	111	786	2511	5	15	51	595	2
Synechococcus CC9605	1421	1485	335	2	64	80	54	2
Prochlorococcus marinus MIT 9515	540	1042	243	4	59	86	120	0
Synechococcus sp WH8102	146	272	35	0	16	29	16	1
Alteromonas macleodii Deep ecotype	10	2	2	426	4	3	5	406
Prochlorococcus marinus phi P-SSM4	179	104	55	0	18	9	4	0
Nitrosopumilus maritimus SCM1	1	2	44	260	0	2	188	1728
Prochlorococcus marinus phi P-SSM2	135	74	51	0	4	3	1	0
Prochlorococcus marinus CCMP1375	19	24	126	3	0	1	58	3
OM42 clade HTCC2255	36	45	50	24	6	8	11	10
Erythrobacter sp. SD-21	69	7	13	55	2	0	2	5
Acinetobacter baumannii SDF	101	9	2	27	0	0	0	2
Prochlorococcus marinus str. MIT 9211 MIT9211	13	25	91	2	0	3	34	0
Tenacibaculum sp. ME0152	35	31	31	17	9	3	12	1
Prochlorococcus marinus MIT9313	2	4	101	1	2	1	21	24
Prochlorococcus marinus MIT 9303	9	6	87	0	1	0	12	2
Synechococcus RCC307	37	31	14	8	1	4	12	0
Synechococcus sp. RS9916 RS9917	28	35	15	2	6	2	12	0
Flavobacteriales bacterium ALC-1	17	23	23	12	3	9	8	1
Kordia algicida OT-1	22	27	14	11	5	2	3	0
Acinetobacter baumannii ACICU	44	2	11	11	0	0	0	1
Rhodospirillales sp. BAL199	15	13	8	33	3	6	10	7
Candidatus Vesicomyosocius okutanii HA	2	4	8	54	0	3	5	7
Pseudomonas syringae phaseolicola 144BA	38	5	4	19	2	1	1	4
Candidatus Ruthia magnifica	2	4	3	51	1	2	6	78
Xanthomonas campestris B100	36	6	3	15	0	0	0	15
marine gamma proteobacterium HTCC2080	24	14	11	7	18	11	8	5
Flavobacteriales sp. SCB49	25	13	9	5	2	5	3	3
Flavobacteriales sp. BAL38	23	12	13	4	1	3	7	2
Synechococcus sp. WH5701	14	14	13	11	2	4	7	3
Staphylococcus aureus phi G1	45	4	1	1	0	0	0	0
Brevundimonas sp. BAL3	27	6	2	16	1	0	2	0

Table S3. Normalized gene expression of *Pelagibacter* strain HTCC7211 (top 60 highly expressed).

ORF number	25m	75m	125m	500m	annotation
1207	19.9	43.3	0.9	0.0	extracellular solute binding protein, family 1
1263	3.2	14.9	1.8	37.8	spermidine/putrescine-binding periplasmic protein
507	22.1	28.6	18.6	1.8	bacteriorhodopsin
244	26.5	2.9	0.0	0.0	protein of unknown function
623	8.8	26.0	2.1	0.0	conserved hypothetical protein
631	22.7	8.7	0.9	0.0	acetaldehyde dehydrogenase II (acdh-ii)
1226	0.0	2.9	18.6	21.6	conserved hypothetical protein
664	0.0	0.0	18.6	0.0	hypothetical protein
1019	3.2	3.9	3.1	16.2	ABC transporter
1094	1.7	13.6	2.3	3.2	Bacterial extracellular solute binding protein, family 7
371	13.2	1.4	1.2	0.0	HeAt shock protein e
1170	13.2	0.0	0.0	0.0	selenium binding protein
914	5.3	13.0	4.1	0.0	ribosomal protein L13
1328	0.0	4.3	12.4	0.0	pilin (bacteria filament)
954	12.4	3.3	0.9	0.0	conserved hypothetical protein
1159	11.8	0.6	0.5	0.5	GTP cyclohydrolase I
243	10.1	11.0	3.2	5.4	Na+/solute symporter (SSF family)
989	0.0	0.0	0.0	10.8	hdig domain protein
544	0.0	0.0	3.1	10.8	tripartite carboxylate transporter, dctp subunit
292	8.8	5.8	5.0	10.8	tripartite carboxylate transporter - dctp subunit
286	1.8	0.0	0.9	10.8	conserved hypothetical protein
195	0.5	0.0	0.0	10.8	chaperone protein DnaJ
823	0.0	1.1	9.3	0.0	transcription termination/antiterminator factor NusG
961	8.8	0.0	2.1	5.4	mttA/Hcf106 family, putative
899	8.8	0.0	0.0	0.0	riboflavin biosynthesis protein RibD
878	8.8	0.0	0.0	0.0	ABC transporter permease component
836	8.8	0.0	0.0	0.0	ribosomal protein L23
743	8.8	8.7	0.0	0.0	conserved hypothetical protein
707	8.8	0.0	0.0	0.0	conserved hypothetical protein
674	8.8	5.8	0.0	1.1	ABC transporter, quaternary amine uptake transporter (QAT) family, substrate-binding protein, putative
595	8.8	0.0	0.0	0.0	conserved hypothetical protein
589	8.8	0.0	0.4	0.0	3-oxoacyl-[acyl-carrier-protein] reductase
472	8.8	0.0	0.0	0.0	modification methylase
377	8.8	0.0	0.0	0.0	conserved hypothetical protein
1337	8.8	5.8	0.0	0.0	type II Secretion Pill
1316	8.8	0.0	0.0	0.0	glutaredoxin 3
1132	8.8	0.0	0.0	0.0	glutathione-dependent formaldehyde-activating, GFA, putative
1126	8.8	0.0	0.0	0.0	sulfide dehydrogenase
920	0.0	8.7	0.0	2.7	6-O-methylguanine DNA methyltransferase
90	2.9	8.7	1.5	0.0	translation initiation factor IF-1
42	0.0	8.7	0.0	0.0	3-oxoacyl-[acyl-carrier-protein] reductase, putative
349	0.0	8.7	0.9	0.0	acid tolerance regulatory protein actr
30	0.0	8.7	0.0	0.0	UDP-glucose 4-epimerase
293	5.9	8.7	0.0	0.0	mannitol transporter
205	0.0	8.7	0.0	0.0	putative porin
1227	0.0	8.7	0.0	0.0	conserved hypothetical protein
1214	7.4	8.7	1.7	8.6	substrate-binding region of ABC-type glycine betaine transport system
1074	4.4	8.7	2.1	0.0	serine--glyoxylate aminotransferase
821	7.9	2.8	4.3	0.9	translation elongation factor Tu
687	7.6	7.8	4.6	2.7	taurine transport system periplasmic protein
1225	4.4	7.6	1.3	3.9	ammonium transporter
420	5.3	7.2	2.8	0.9	ATP synthase subunit C, putative
1129	6.6	7.1	7.0	3.6	non-specific DNA binding protein HBsu
737	7.1	1.7	0.0	0.0	tripartite carboxylate transporter- dctp subunit
1269	2.2	2.9	0.4	7.0	ABC proline/glycine betaine transporter, periplasmic substrate-binding protein
855	6.6	2.5	0.0	0.0	ribosomal protein S13p/S18e
480	6.6	0.0	0.0	0.5	cell division protein FtsZ
96	0.0	0.0	6.2	0.0	ribosomal protein L34
815	0.0	0.0	6.2	0.0	prepilin-type N-terminal cleavage/methylation domain protein
637	0.0	0.0	6.2	0.9	molybdenum cofactor biosynthesis protein C

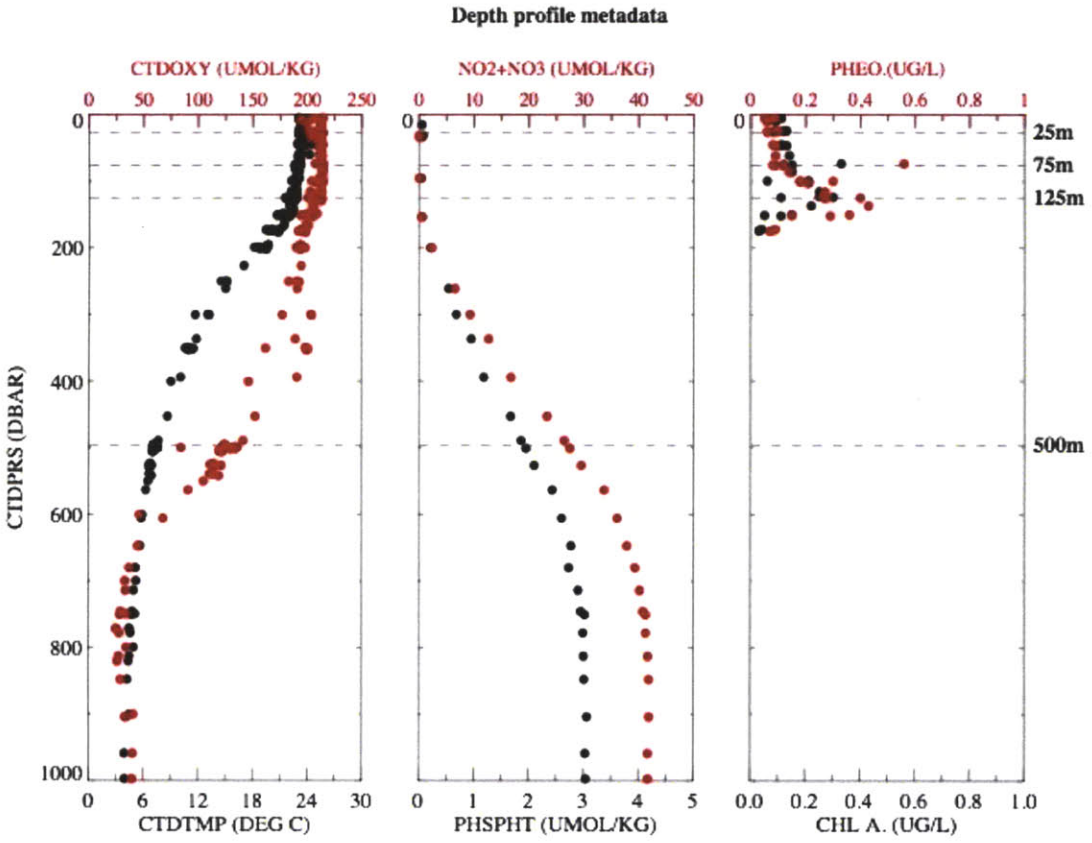


Figure S1. Biogeochemical data of the sampling station collected on the cruise. Dashed lines indicate four sampling depths. Data source: http://hahana.soest.hawaii.edu/hot/hot-dogs/interface.html.

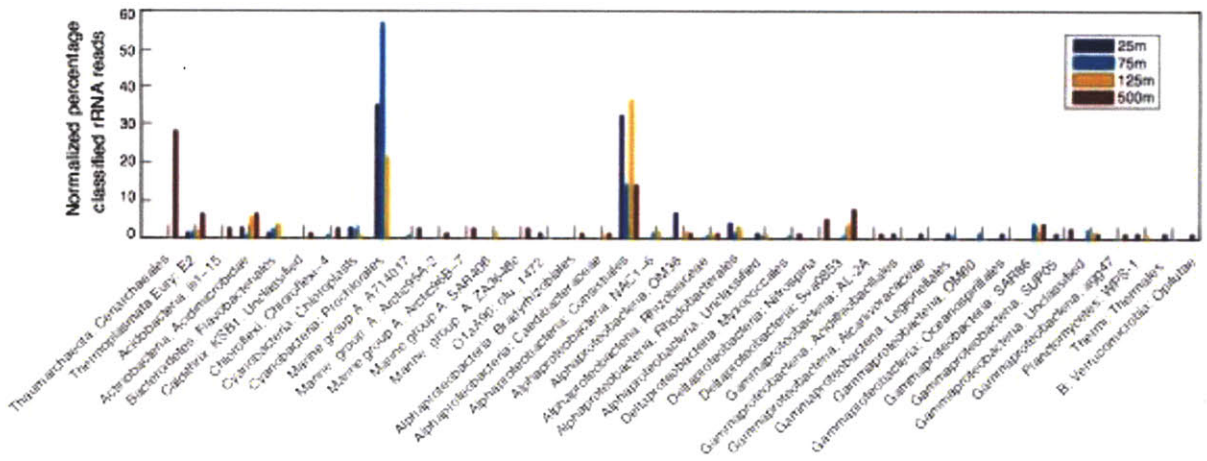


Figure S2. Taxonomic classification based on 16S rRNA-bearing shotgun sequences. The shotgun libraries and pyrosequencing libraries were constructed from identical DNA samples. Taxonomic assignments were binned at the Order level, using the Hugenholtz taxonomy of Greengenes (see Supplementary Methods). 16S rRNA sequences that could not be classified were excluded from the analysis. Y-axis scale represents the percentage of the total classified 16S rRNA reads. Only taxa that represented ≥ 1% of all classified reads are displayed.

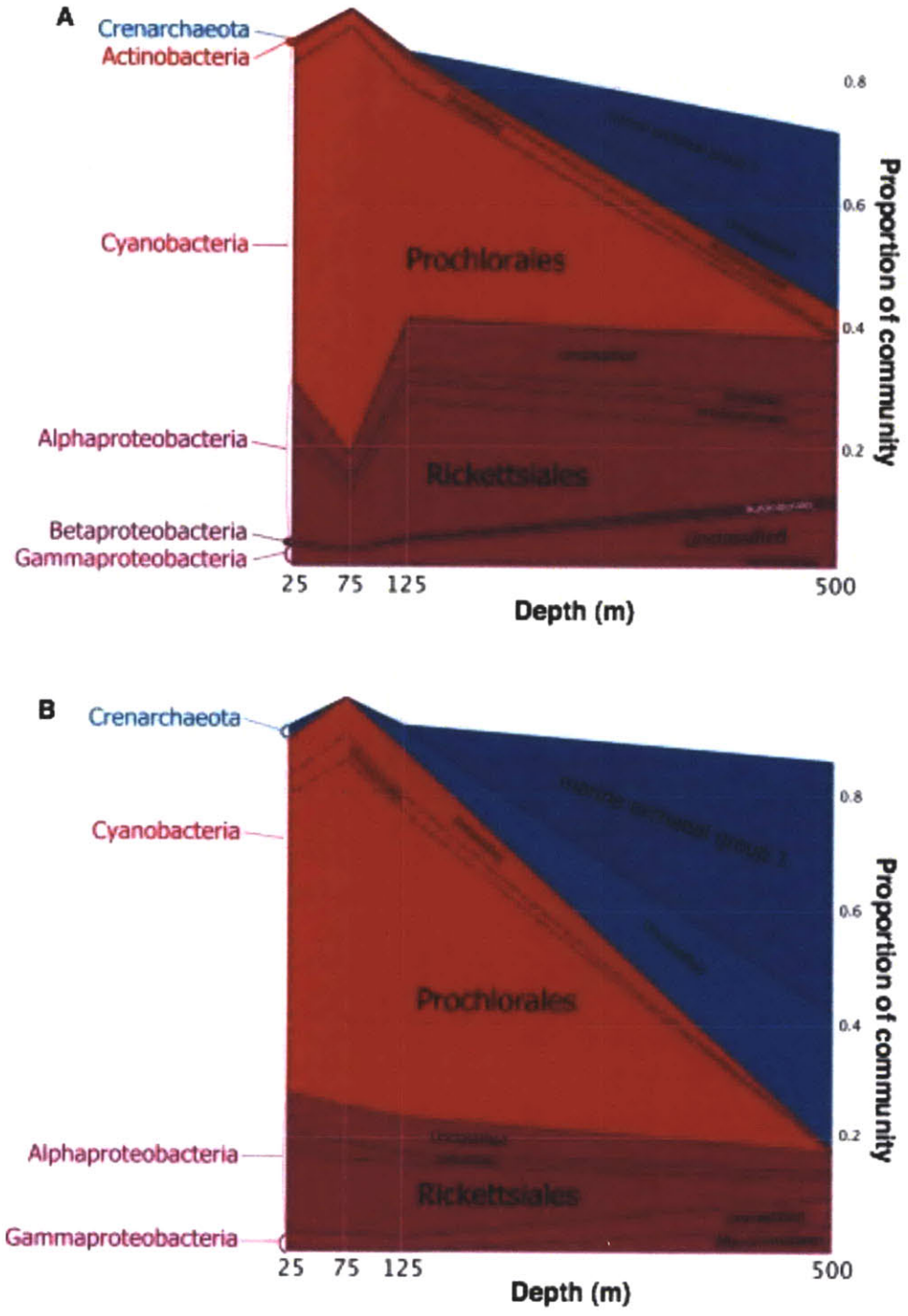


Figure S3. Stacked area plot showing taxonomic classification of protein-coding sequences. Taxonomic assignments were based on BLASTx against NCBI-nr protein database, using MEGAN (Huson et al., 2007), with default settings. Upper panel represents DNA samples, and lower panel represents cDNA samples.

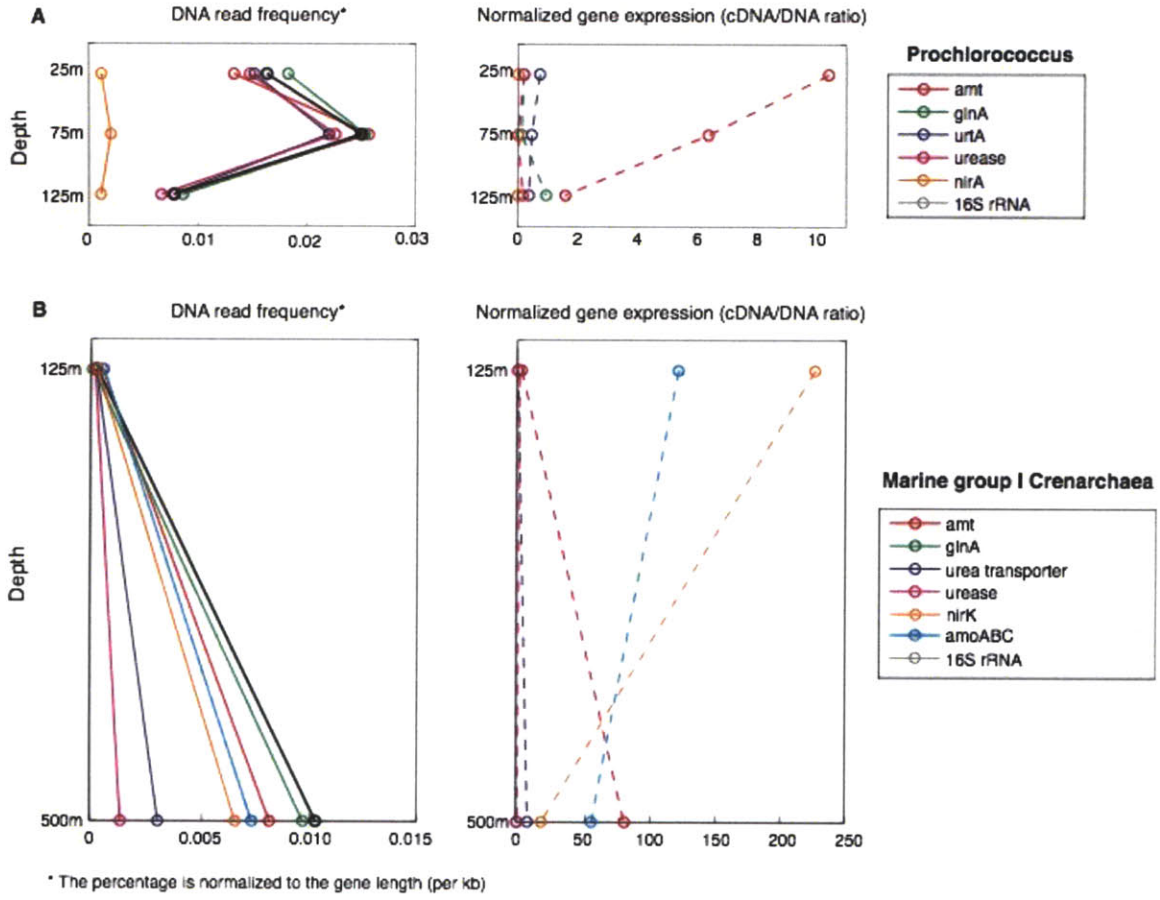


Figure S4. Abundance and normalized expression levels of genes involved in nitrogen metabolism. The abundance of 16S rRNA genes was used to indicate taxon abundance, and was compared to detected abundance of a suite of functional genes (listed in figure legends). Normalized gene expression was calculated as described in Supplementary Methods. (A) Functional genes putatively originated from *Prochlorococcus* populations, in the three euphotic zone samples. (B) Functional genes putatively originated from marine group I crenarchaeota populations in the deep euphotic zone and the mesopelagic samples.

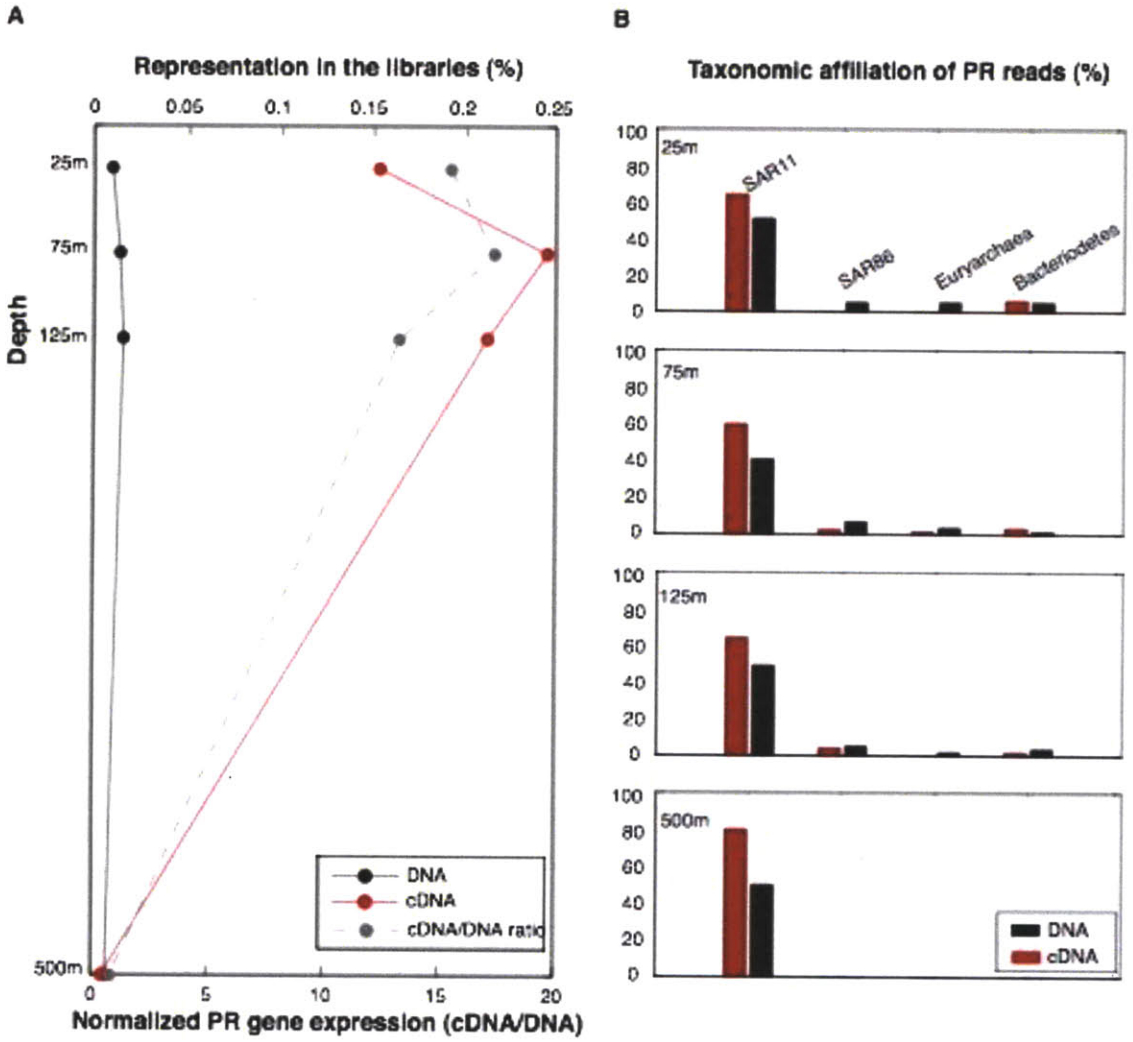


Figure S5. Abundance, expression and taxonomic origins of Proteorhodopsin (PR)-encoding reads. (A). Representation of PR-encoding reads in the DNA and cDNA data sets, and their normalized expression levels in the four depths. (B) Putative taxonomic breakdown of PR sequence reads. PR sequences were first identified by BLASTx against NCBI-nr database, then aligned to a custom PR sequence database (McCarren & DeLong, 2007), and finally added to the backbone PR phylogenetic tree using ARB's "parsimony insertion" feature. The taxonomic origin of a PR-encoding sequence was assumed the same as that of the most related sequence in the PR phylogenetic tree.

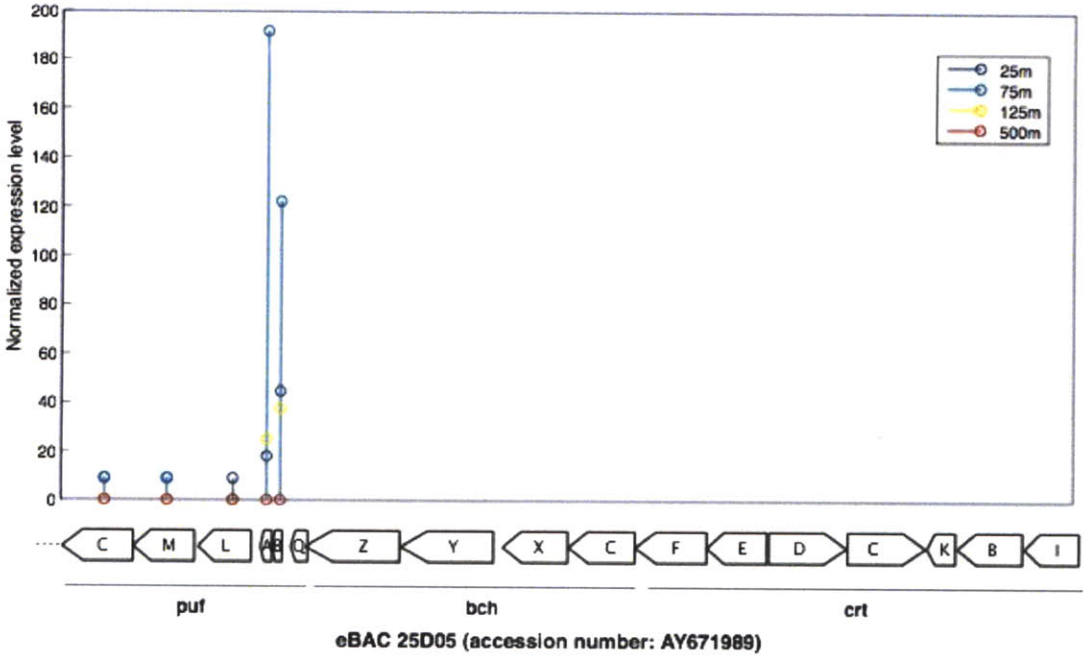


Figure S6. Expression of genes involved in aerobic anoxygenic phototrophy (AAP), using a *Roseobacter*-like BAC clone insert as a reference. The BAC clone is eBACred25D05 with an accession number of AY671989. *puf*: light-harvesting and reaction center genes; *bch*: bacteriochlorophyll biosynthesis genes; *crt*, carotenoid biosynthesis genes. Y-axis scale represents normalized cDNA to DNA ratio (normalized expression level; see Supplementary Methods).

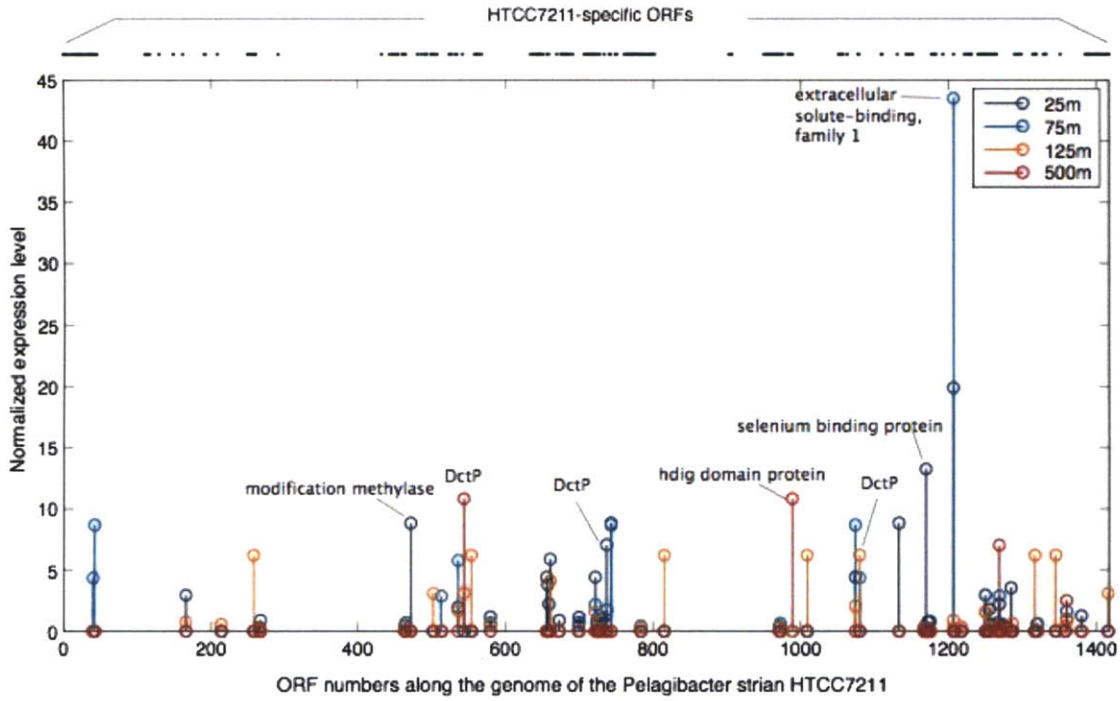


Figure S7. Gene expression of *Pelagibacter* HTCC7211-specific ORFs. The HTCC7211-specific ORFs are denoted by the black dots on top the panel, and were identified as ORFs lack of apparent homology to ORFs in the two coastal *Pelagibacter* strains HTCC1062 and HTCC1002 (see Supplementary Methods).

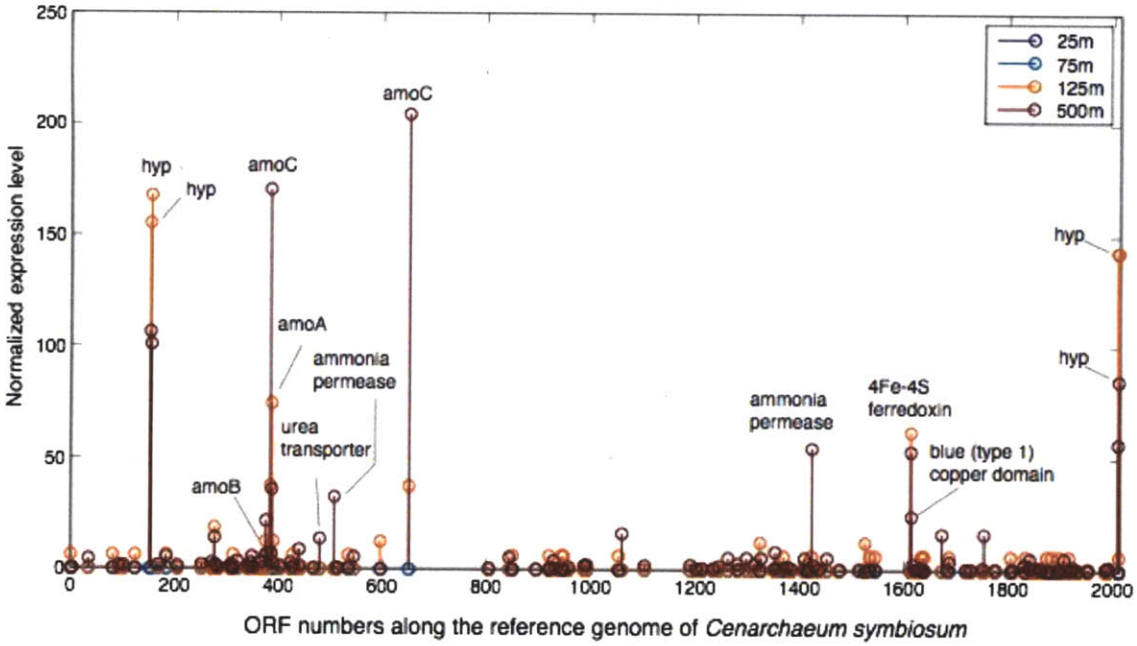


Figure S8. Genome-wide expression profiles of marine crenarchaea-related populations, in all four depths. The x-axis, y-axis, and figure legend are the same as those in Figure 5.

CHAPTER FOUR

Experimental metatranscriptomics: probing microbial transcriptional responses to simulated upwelling in the open ocean

Yanmei Shi, Jay McCarren, Edward F. DeLong

This chapter is the outcome of a collaborative effort with Jay McCarren, composed of three in-parallel yet independent experiments. The first experiment led by Jay McCarren has been published (McCarren *et al*, *Proc. Natl. Acad. Sci. USA.* **107** 16420-16427 (2010)). Some of the methods described in this chapter overlap with those in McCarren *et al* (2010). Corresponding supplementary information is appended.

Chapter 4: Experimental metatranscriptomics: probing microbial transcriptional responses to simulated upwelling in the open ocean

Abstract

Deep water mixing events in the open ocean provide a periodic yet significant source of inorganic nutrients to the nutrient-limiting surface waters, often causing (large cell) phytoplankton blooms and consequently impacting carbon cycles. Here we set out to understand how surface microbial assemblages respond, at the molecular level, to a simulated deep sea water (DSW) mixing experiment. Flow cytometric and transcriptomic analysis both revealed apparent growth response of an *Alteromonas*-like population in the DSW-amended treatment from 12 hr and onward, of which chemotaxis, cell motility, and carbon metabolism pathways were significantly up-regulated. Other major taxonomic components of the community were relatively unresponsive with respect to cell abundance, but changes in genome-wide transcriptional activities were readily detectable. As the dominant phytoplankton in the initial water sample, *Prochlorococcus* showed significantly elevated gene expression level for carbon fixation-related genes and some photosynthesis genes, as well as increased cell density, relative to the control. Captured cyanophage DNA and cDNA profiles resembled possible transition from phage pseudolysogeny to active lysis. These observations suggested that previously reported phytoplankton shift from *Prochlorococcus* to larger cells might not be due to decrease in *Prochlorococcus* cellular fitness but more likely caused by higher grazing and/or phage-induced mortality rate. Finally, we compared DNA and cDNA reads of DSW-responsive *Alteromonas* and those of dissolved organic matter (DOM)-responsive *Alteromonas*, reported by McCarren et al (McCarren et al., 2010). A set of genes showed differential abundance between these two *Alteromonas* populations, majority of which were transposable and phage-related genes. Additionally, specific KEGG pathways recruited significantly different numbers of transcripts between the two *Alteromonas* populations from the two different treatments, suggesting perturbation-specific metabolic responses. In total, our study demonstrates the power of experimental metatranscriptomics to reveal microbial dynamics and interactions under specific environmental influences, at a higher resolution and on a finer time scale.

Introduction

Metatranscriptomic surveys have provided useful information about the composition of microbial transcriptomes in natural samples at times of sampling (Frias-Lopez et al., 2008). Comparative analyses have further revealed differential transcriptional activities for samples across geochemical gradients (Chapter 3 of this thesis) (Hewson et al., 2010; Poretsky et al., 2009). However, it is poorly understood to what extent such variations are neutral or reflect microbial responses to environmental cues, since it is difficult to deconvolute complex biogeochemical dynamics characterizing each environment.

To that end, the application of metatranscriptomics in experimental settings such as laboratory microcosms and field mesocosms (termed experimental metatranscriptomics hereafter) can facilitate more controlled assessment of community transcriptional responses to environmental changes over time. The environmental variation examined can be natural (for example, tracking changes in gene expression as a function of the daily cycle) or applied (for example, monitoring changes in gene expression following changes to nutrient levels). Recently, McCarren *et al* conducted a microcosm experiment where high molecular weight dissolved organic matter (HMWDOM) was added to a seawater sample, and microbial community transcriptomes were sampled and sequenced over the course of 27 hours (McCarren et al., 2010). The data revealed an apparent successional community response and transcriptional changes, suggesting specific resource partitioning of DOM by different bacteria species. This molecular-level resolution complements significantly to conventional bulk measurements such as community substrate incorporation and respiration in incubation experiments (Carlson et al., 2004; McAndrew et al., 2007).

In tandem with the HMWDOM amendment, we carried out deep sea water (DSW) amendment, hoping to reproduce some of the microbial responses and dynamics induced by deep mixing/nutrient loading events. Nutrient availability is central to microbial activity and thus essential to all energy and matter fluxes mediated by microbes (Arrigo, 2005; Karl, 2007). Nearly 80% of the surface waters of the global ocean are considered nutrient-limiting, characterized by low rates of new production and export (Longhurst, 1998). In contrast, nutrient concentration increases sharply in deep waters due to net release from and oxidation of exported organic matter (Karl, 2002). At our study site in the North Pacific subtropical gyre (NPSG), the nitrate concentration at 1000 m depth is approximately 42 μM, but is generally < 5 nM in the upper 100 m.

Nutrient repletion/depletion experiments on cultivated isolates under laboratory settings have yielded valuable information on microbial phenotypic responses and the corresponding genetic basis for these responses (Konneke et al., 2005; Lindell et al., 2002; Moore et al., 2002). However, the pure compound nutrient additions (such as nitrate, phosphate, and glucose), frequently done in such experiments (Carlson et al., 2002; Karl et al., 2008), may not represent the environmentally relevant nutrient loading. This is because 1) some limiting nutrients may

remain unidentified, and 2) nutrient co-limitation is a recurrent scenario in the open ocean (Arrigo, 2005; Aumont, Maier-Reimer, Blain & Monfray, 2003; Saito, Goepfert & Ritt, 2008). In this study, we mimicked nutrient loading in seasonal deep mixing and eddy diffusive processes, by adding 2L of 700 m water to 18L of 75 m surface water sample. The responses and kinetics of microbial community structure and transcriptional changes were then monitored using integrated metagenomic and metatranscriptomic approach.

It has been shown in multiple studies that the addition of nutrient rich deep waters to nutrient depleted surface waters stimulates primary production significantly over the course of days to weeks (Carlson et al., 2004; McAndrew et al., 2007). In addition, phytoplankton community structure seems to change in favor of larger sized phototrophs such as diatoms and *Trichodesmium*. However, very little is known about how picoplanktons and heterotrophic bacterioplankton community reacts to the nutrient addition. Phylogenetic analyses of time-series samples have identified some taxa that appear to increase in numbers over days or weeks after deep-water mixing events (Hansell & Carlson, 2001; Morris et al., 2005), but the short-term molecular-level responses of microbial populations to deep water mixing events remain uncertain.

Finally, nutrient addition has been shown to significantly affect production of new DOM and consumption of seasonally accumulated DOM (Carlson et al., 2002; Hansell & Carlson, 2001). The microcosm experiments employing HMWDOM treatment (McCarren et al., 2010) and DSW treatment (reported here) were performed in parallel. Here, we compared the community transcriptomic dynamics in these two treatments, in an effort to gain insights into microbial processes relevant to the intercorrelated effects of nutrient and organic carbon cycling dynamics.

Methods

Experimental setup and sample collection

Seawater for microcosm incubation experiments was collected (23°12.88′N, 159°8.17′W) from the 75 m depth, predawn, on August 16, 2007, during the Center for Microbial Oceanography: Research and Education (C-MORE) BLOOMER cruise

(http://hahana.soest.hawaii.edu/cmorebloomer/cmorebloomer.html). Deep sea water was collected from 700-m depth, in the oxygen minimum zone region, equilibrated to surface water temperature, and added to a 75-m water sample at a 1:9 ratio. The depth of water samples and the mixing ratio were chosen such that they are consistent with those in a previously reported experiment performed at the same site (McAndrew et al., 2007). The experimental design and sampling strategy for DSW and HMWDOM amendments are illustrated in Figure 1. See **Supplementary Information** for details on the seawater collection and microcosm preparation.

Flow Cytometry and Cell Sorting

At each time point, 1 mL of seawater was preserved with 0.125% glutaraldehyde (final concentration), frozen in liquid nitrogen, and stored at –80 °C for subsequent flow cytometric analysis and cell sorting using an Influx (Becton Dickinson). Before counting and sorting, samples were stained with SYBR Green (Invitrogen) for 15 min, and DNA-containing cells were identified based on fluorescence and scatter signals (Marie, Partensky, Jacquet & Vaulot, 1997). See **Supplementary Information** for further details on cell sorting and rRNA amplicon sequencing from the sorted population.

Ribosomal RNA (rRNA) subtraction, RNA amplification, cDNA synthesis, and pyrosequencing

Subtractive hybridization using sample-specific biotinylated rRNA probes was used to remove bacterial 16S and 23S rRNA molecules from total RNA samples, as previously described by Stewart *et al* (Stewart et al., 2010). Subtracted RNA was amplified, cDNA synthesized, and pyrosequenced as previously described (Frias-Lopez et al., 2008) with minor modifications. See **Supplementary Information** for more detail.

Bioinformatics analysis

The 3 DNA and 10 cDNA data sets included roughly 5 million FLX reads, with an average read length of 200 bp. Low-quality and exact replicate reads were removed from DNA data using a custom perl script and CD-HIT (Li & Godzik, 2006); rRNA reads were identified as described (Chapter 3; Shi *et al*, 2010, in press), and removed from cDNA reads. Non-rRNA sequences were compared to NCBI-nr, and KEGG databases using BLASTX for functional analyses. Taxonomic analysis and functional gene analysis were described in detail in the

Supplementary Information. A custom microbial genome database (ORF amino acid) was constructed from 2067 publicly available microbial genome sequences (as of January 2009), and was used to recruit cDNA and DNA reads. See **Supplementary Information** for further details

Results and Discussion

Nutrient loading in the DSW-treatment

Microcosm incubation experiments are conventionally carried out over the course of days to weeks, in order to capture bulk level dynamics of the microbial community (Braddock, Ruth, Catterall, Walworth & McCarthy, 1997; Carlson et al., 2004). However, this has potential to amplify artifacts due to "bottle effects", that drastically change microbial community profiles and activities simply due to confinement (Fuhrman & Azam, 1980; Williams, 1981). At the molecular level, microorganisms respond to external perturbations on the time scale of minutes to hours (Kort, Keijser, Caspers, Schuren & Montijn, 2008; Lindell et al., 2007; Steglich et al., 2010). In addition, microbes in the oligotrophic open ocean generally grow with turnover times between 1 to 25 days (Whitman et al., 1998). For these reasons, experimental metatranscriptomics provide a desirable platform to capture microbial gene expression dynamics in microcosm experiments with short incubation times, minimizing potential bottle effects and significant community structure change.

The experiment was carried out in the summer time (August, 2007), when the water column at Station ALOHA usually is highly stratified and nutrient-depleted (Dore, Letelier, Church, Lukas & Karl, 2008). By mixing 10% 700-m depth water sample with 90% 75-m depth sample, we added roughly 700 X ambient concentration of inorganic nitrogen, 4 X inorganic phosphorus, and 3.4 X silicate (Supplementary Table S1). These saturating concentrations allow maximal uptake of nitrogen and phosphorus for most oligotrophic cells. In addition, nutrients such as inorganic carbon, iron and other trace metals were also enriched in the DSW treatment relative to the control, but we do not have quantitative measurements of their concentrations.

DSW-induced cell dynamics

Microbial cell counts remained constant in the control microcosm throughout the 27

hours, while the DSW treatment microcosm showed a slight yet clear increase in total cell counts (Supplementary Table S2). Flow cytometric enumeration indicated that the majority of this increase in cells was attributable to the growth response of a specific population of larger, high-DNA-content cells (Figure 2A), which were later separated for further analyses. These large, high-DNA-content cells were isolated and collected via fluorescence-activated cell sorting and used to generate a SSU rRNA gene PCR library for sequencing. Near full-length rRNA gene sequences (9 sequences in total) from the sorted cells from DSW-amended sample recovered were all affiliated with the *Alteromonas macleodii*. In comparison, 5 out of 11 rRNA sequences from HMWDOM-amended sample fell into the *Alteromonas* clade, while others belonged to *Methylophaga*, and *Rhodobacteraceae* ((McCarren et al., 2010); Supplementary Figure S1).

A. macleodii is a ubiquitous marine heterotrophic gamma-proteobacterium, that is readily culturable but usually present in low abundance in the open ocean (DeLong et al., 2006; Eilers, Pernthaler, Glöckner & Amann, 2000). Isolates from the open ocean can be clustered into two major genotypic groups or ecotypes, surface and deep water, by multi-locus sequence analysis and comparative genomic analysis (Ivars-Martinez et al., 2008; Ivars-Martínez et al., 2008; López-López, Bartual, Stal, Onyshchenko & Rodríguez-Valera, 2005). Here, phylogenetic reconstruction with 29 full length 16S rRNA sequences of *A. macleodii* isolates (exported from the Silva SSU rRNA database as of October 2010), clustered all *A. macleodii* 16S rRNA amplicons with the surface ecotype isolates, except one sequence from HMWDOM treatment that clustered with the deep ecotype isolates (Figure 2B). Since *A. macleodii* are common responders to perturbation experiments (Schäfer, Servais & Muyzer, 2000; Zemb, West, Bourrain, Godon & Lebaron, 2010), as seen here in both DSW and DOM treatments, we then explored the potential genomic and gene expression differences of responsive *A. macleodii* between the two treatments (see below in the section of *Alteromonas*-centric analysis).

Taxonomic composition change over the course of incubation

Due to limited water volume in the microcosm experiment, we collected three community genomic DNA samples (T0, 0 hr; Control T5, 27 hr; and DSW T5, 27 hr), which were pyrosequenced on the Roche 454 FLX platform, yielding ~420,000 to 550,000 reads per sample (Table 1). Roughly 0.4% of these reads were designated SSU rRNA sequences, allowing taxonomic classification using Greengenes classification tools (Figure 3A). The most significant

taxon change in the DSW treatment appeared to be the relative increase of gamma-proteobacteria, especially the genus *Alteromonas*, from < 1% in the T0 DNA sample to > 11% in the DSW T5 sample. The community structure in the Control microcosm, for both T0 and T5, appeared very similar to typical taxonomic profiles recovered from the same depth at Station ALOHA (DeLong et al., 2006; Frias-Lopez et al., 2008; McCarren et al., 2010). This observation supports our initial assumption that the microbial community would not change significantly over 1 day of incubation, allowing detection of taxon-specific changes in transcript abundance, without a normalization to corresponding gene abundance, as we routinely apply in metatranscriptomic survey studies (Chapter 2 and Chapter 3). Although, it is worth pointing out that, drastic community structure changes have been observed in other perturbation experiments (McCarren et al., 2010), which complicates the assessment of gene expression changes for populations showing very different abundance in control and treatment samples.

Taxonomic classification of putative protein-coding sequences (sequences that have a significant match to NCBI-nr protein database) in the three DNA data sets generally paralleled the patterns observed for rRNA gene taxon abundance (Figure 3B). The main difference was the detection of phage related sequences, which was not possible for rRNA gene-based analysis. Although the relative abundance of cyanobacterial-like sequences was equivalent for Control T5 and DSW T5, the cyanophage relative abundance was apparently lower in the DSW treatment (Supplementary Figure S2). If we assume that phage DNA captured by our sampling method (see Supplementary Methods) originated from infected host cells, as previously hypothesized (DeLong et al., 2006), our observation suggested a smaller fraction of infected cyanobacterial cells in the DSW treatment microcosm. As a comparison, such decrease in phage sequences was not observed in the DOM T5 sample (data not shown).

Community transcriptome dynamics

Community RNA samples at each time point were sequenced, and reads mapping KEGG categories provided an overview of the functional processes driving transcriptional differences between the DSW and control samples. To examine the overall relatedness of the 10 community transcriptomes, we clustered the RNA datasets based on the distribution of reads matching KEGG gene categories (KEGG 3 hierarchy level; Figure 4). A general pattern was apparent from the analysis: the community transcriptome dynamics is affected by at least two factors: time

effect and treatment effect. All T1 and T2 samples, including control and DSW treatment, clustered together to the exclusion of all T3, T4, and T5 samples. Within these two major clusters, treatment effect was obvious, as control and treatment samples formed clear sub-clusters (Figure 4).

Reads mapping to NCBI-nr functional genes with taxonomic affiliations allowed us to group the putative protein-coding genes differentially represented between the DSW and Control transcriptomes (identified using the R package DEGseq (Wang, Feng, Wang, Wang & Zhang, 2010)), into representative taxa (7 *Prochlorococcus* strains, 3 *Pelagibacter* strains, and 2 *Alteromonas* strains). A total of 1296 NCBI-nr reference genes were designated to be more actively expressed in DSW treatment relative to the control, ~ 42% - 65% of which were categorized as one of the 12 reference strains (Supplementary Figure S2, upper panel). On the other hand, a total of 1578 NCBI-nr reference genes were found proportionally under-represented in the treatment transcriptomes, ~ 76% - 88% of which belonged to the 12 reference taxa (Supplementary Figure S2, lower panel). In addition, several other general patterns were evident from this analysis. First, *A. macleodii* like transcripts dominated DSW-enriched transcriptomes from T3 onwards, consistent with an increase in *Alteromonas* population cell number (Figure 2). Specifically, 34-81 NCBI-nr genes related to *A. macleodii* 'Deep ecotype' (referred to AltDE hereafter), and 47-102 NCBI-nr genes related to *A. macleodii* ATCC 27126 (surface ecotype, referred to AltATCC hereafter) were found enriched in at least one of DSW treatment samples. Most of these *A. macleodii* transcripts fell into a handful of functional processes including chemotaxis and flagellar biosynthesis (see below). Next, *Prochlorococcus* like transcripts comprised a large fraction of both the DSW-enriched and DSW-depleted transcript classes. This in part may suggest *Prochlorococcus* cells up-regulate as many functional pathways as they down-regulate in responsive to DSW amendment, but might simply reflect issues related to the relative quantification. Specifically, the increased abundance of some transcripts may cause apparent decrease in others whose absolute abundance may not have changed. Finally, *Pelagibacter* demonstrated relatively lower transcript abundance in the DSW treatment, which again could be attributed to the proportionally higher abundance of *Alteromonas* transcripts. A discussion on the merits and complications of relative versus absolute quantification (Gifford et al., 2010) of meta-omics studies is beyond the scope of this dissertation, but it is important to bear such caveats in mind when drawing conclusions from

meta-omics data sets.

Taxon-specific responses to DSW addition, inferred from genome-centric transcriptome analyses

In this simulated deep mixing event, autotrophic, heterotrophic, and bacteriophage populations displayed distinct shifts in their relative transcript abundance, compared to their counterparts in the control samples. To examine taxon-specific transcriptional responses, a custom reference database was constructed from 2067 publicly available microbial genome sequences (fully sequenced and draft genomes as of January 2009, plus several extra draft genomes). Seven populations showed discernible fold changes in their relative representation (Figure 5). *A. macleodii* deep ecotype and surface ecotype (Ivars-Martinez et al., 2008), low-light *Prochlorococcus* eNATL ecotype (Coleman & Chisholm, 2007), and cyanophage P-SSP7 (Lindell et al., 2004) displayed an elevated, genome-wide transcript abundance. In contrast, high-light *Prochlorococcus* eMIT9312 and eMED4 ecotypes, as well as *Pelagibacter* strains, showed relatively lower genome-wide transcript abundance (Figure 5).

We further examined up-regulated and down-regulated genes within specific genomes. This genome-centric analysis differs from the community-level analysis (Supplementary Figure S2): in the latter, transcriptional changes in one taxon may potentially affect another taxon, whereas here differentially expressed genes in one genome were identified by comparing only transcripts recruited to that specific genome.

Alteromonas: Alteromonas (in this study mostly *A. macleodii*), is known as r-strategist with preference for the nutrient rich micro-niche (Acinas, Anton & Rodriguez-Valera, 1999; López-López et al., 2005). It is also commonly found to respond rapidly to environmental perturbations such as transient nutrients (Cappello et al., 2007; Zemb et al., 2010). Significantly enriched transcripts in the DSW treatment included those involved in chemotaxis and cell mobility (Figure 6), underlining the chemotactic nature of this gamma-proteobacterium. Additionally, key genes required for the glutamine synthetase/glutamate synthase (GS/GOGAT) cycle involved in nitrogen metabolism and amino acid synthesis, as well as genes involved in citric acid cycle and gluconeogenesis were also up-regulated in the treatment. Finally, substrate transport and protein synthesis appeared to be more abundant in the DSW-amended population as well (Figure 6).

Seymour *et al* have experimentally demonstrated strong and rapid chemotactic responses of three open-ocean proteobacterial strains to the extracellular products of cyanobacteria *Prochlorococcus* and *Synechococcus* (Seymour, Ahmed, Durham & Stocker, 2010). It is plausible here that the amendment of deep water stimulated extracellular exudation or cell lysis of cyanobacteria, resulting in increased amounts of fresh, labile dissolved organic carbon (DOC) that serves as chemoattractants for *Alteromonas*. This may also be related to responses we observed in *Prochlorococcus* and cyanophages (see below).

Comparison of DSW-responsive Alteromonas and DOM-responsive Alteromonas: To better understand the genomic and transcriptomic differences between the *Alteromonas* populations that responded to DSW amendment and those that responded to HMWDOM amendment, we examined those datasets more closely for differences. Do the responsive *Alteromonas* cells in the different treatments represent genomically coherent populations? Do they employ similar metabolic strategies to respond to the two different environmental perturbations? To address these related questions, we first pulled out all reads in T5 samples that were assigned to *Alteromonas* (Supplementary Table S3), and compared their nucleotide-level similarity, gene content, and transcript abundance, based on *A. macleodii* reference genomes (AltATCC and AltDE), that represent two different ecotypes of this species (Ivars-Martinez et al., 2008). Several general patterns arose from our comparative analyses.

First, the majority of *Alteromonas* cells in both treatments (HMWDOM, McCarren et al; DSW, this work) were more closely related to the surface ecotype, as revealed by nucleotide diversity analysis of sequence reads mapping to the reference genomes. The *Alteromonas*-like reads were dominated by genotypes sharing ~98% nucleotide identity with AltATCC (Supplementary Figure S3, upper panel), and sharing ~81% nucleotide identity with AltDE (Supplementary Figure S3, lower panel). AltDE-like genotype was also detected, though at a much lower abundance (a smaller peak around 81% nucleotide identity against AltATCC genome). In addition, this AltDE-like genotype appeared to constitute a larger fraction of *Alteromonas* populations in the DOM amendment sample, consistent with the identification of deep ecotype-related 16S rRNA genes amplified from flow-sorted DOM-responsive populations (Figure 2). It is possible that these AltDE-like cells can more readily degrade the relative recalcitrant fraction of added HMWDOM, compared to their surface ecotype counterparts (Ivars-

Martinez et al., 2008).

Second, to examine gene content of DSW- and DOM-responsive *Alteromonas* cells, we compared read frequencies of ORFs derived from the AltDE and AltATCC genomes. The ORFs were divided into three categories: shared by both genomes, AltDE-specific, and AltATCC-specific (see Supplementary Methods). We detected significantly different read frequencies for only 11 ORFs (Supplementary Figure S4), 7 of which were AltDE-specific, almost exclusively transposable and phage-related genes that are typical of mesopelagic Alteromonas populations (Ivars-Martinez et al., 2008). This presumably reflects the higher abundance of transposases in the *Alteromonas* cells in the deep, a feature that seems typical of deep-sea bacteria in general (DeLong et al., 2006). Deep mixing events in nature mix not only chemical compounds but also microbial assemblages, creating perturbed environments for both surface and mesopelagic microbial communities. Carlson *et al* showed in a simulated deep mixing experiment that mesopelagic heterotrophic microbes can readily degrade semilabile DOC produced in the surface water (Carlson et al., 2004), raising the possibility that *Alteromonas* cells added from 700-m depth to the microcosm may benefit from exposure to a higher DOC concentration.

Alteromons transcriptomes shared some, but differed in other transcript abundance patterns in response to the DSW- and DOM-amendments. Two component systems were highly expressed in both cases. Chemotaxis, cell motility, and cell growth related genes were particularly abundant in the DSW-amendment transcriptomes. Fatty acid catabolism and downstream carbon metabolism was enriched in the DOM-amendment transcriptomes, suggestive of differential metabolic responses to the carbon contained in the HMWDOM treatment.

Prochlorococcus: *Prochlorococcus*, the smallest known oxygenic phototroph, numerically dominates microbial assemblages in the photic zone of many oceanic regions including our study site (Malmstrom et al., 2010). For this reason, *Prochlorococcus* transcriptomes were well represented in both the control and treatment data sets, and thus changes in genome-wide transcriptional activities were readily detected. Out of 1926 protein-coding genes in the AS9601 genome, transcripts from 1499 genes were detected, 242 of which were designated as differentially expressed using DEGseq (Supplementary Figure S5).

The strongest signal in the DSW amendment samples was the up-regulation of genes

involved in carbon fixation (i.e., genes encoding Rubisco subunits, phosphoglycerate kinase, glyceraldehyde 3-phosphate dehydrogenase, and carboxysome shell protein CsoS1) (Supplementary Figure S6). For instance, the read number of Rubisco large subunit transcripts increased from 200-700 copies in the control samples to 2000-7000 copies in the treatment (~10 fold increase at each time point). On the other hand, one gene showed the strongest down-regulation in the treatment (Supplementary Figure S5), the hypothetical gene A9601_11371.

McAndrew and colleagues (McAndrew et al., 2007) showed in a similar microcosm experiment that, *Prochlorococcus* cell abundance declined significantly (> 70%) after 72 hours in both control and the treatment. Here, over the first 27 hours, *Prochlorococcus* decreased in abundance by 10% in the control and 4.3% in the treatment (Supplementary Table S2). These observations potentially suggested an adverse effect of microcosm incubation on naturally-occuring *Prochlorococcus* cells, which may be temporally alleviated by the addition of deep water (with nutrients replenishment). Additionally, McAndrew *et al* reported that in the treatment that the phytoplankton community shifted from small (< 2 μm diameter) to large (> 10 μm diameter), chl *c* containing and Si utilizing cells. This larger phytoplankton growth response usually occurs after > 2 days of incubation ((McAndrew et al., 2007); Angelicque White, personal communication). Our short-term incubation suggested that prior to the community shift, there appeared to be an initial increase in phototrophic and carbon fixation activity for *Prochlorococcus* populations.

Cyanophages: Phage-mediated microbial mortality is an important component of the microbial food web and thus has fundamental importance in marine carbon and nutrient cycling (Sullivan, Waterbury & Chisholm, 2003; Suttle, 1994; Suttle, 2005). Our sampling method was not intended for capturing free-living phages. Nevertheless, we observed differential gene abundance and transcript abundance in the treatment versus the control for T7-like and T4-like cyanophages, which were presumably derived from the cytoplasm of infected cyanobacterial cells (DeLong et al., 2006). The cyanophage-like sequences were identified using tblastx and blastn with a stringent cutoff, instead of blastx (Supplementary Figure S7). Four T7- and T4-like cyanophage genomes (particularly podoviridae P-SSP7) recruited apparently more cDNA reads and fewer DNA reads in the DSW-amended sample, resulting in higher gene expression ratio for these phage genomes (Figure 8; Supplementary Figure S8). P-SSP7 genes enriched in the DSW

treatment included T7-like RNA polymerase, ribonucleotide reductase, T7-like capsid, T7-like ssDNA binding protein, and possible endonuclease (data not shown).

If we assume that phages were sampled as part of infected host cells, higher phage gene expression and lower phage DNA abundance indicated active lytic processes in the DSW amendment sample, which might provide organic carbon source for co-existing heterotrophs (see discussion in the *Alteromonas* section). This nutrient-induced phage lysis might reflect a type of phage-host interaction state termed pseudolysogeny, a less-understood phenomenon where starved bacterial cells coexist in an unstable relationship with infecting viral genomes (Weinbauer, 2004). Upon nutrient replenishment, the pseudolysogens resolve into either true lysogeny or active production of virions (lysis) (Ripp & Miller, 1997; Ripp & Miller, 1998; Williamson, McLaughlin & Paul, 2001). For this reason, pseudolysogeny effectively supports long-term survival of viruses in unfavourable environments, and therefore has potential ecological significance in surface ocean waters where nutrients are chronically limiting.

Conclusions and future direction

Deep water mixing events in many oceanic regions contribute to phytoplankton blooms, microbial community structure shifts, increases in primary production and secondary bacterial production, that together result in increased levels of carbon cycling (Carlson, Ducklow, Hansell & Smith Jr, 1998; Karl & Letelier, 2008; Lindell & Post, 1995). However, very little is known about the details of how microbial assemblages respond in the early stages of nutrient injections to alter gene expression and metabolic pathways. In this chapter, we used experimental metatranscriptomics to ask how microbes respond transcriptionally to those specific environmental perturbations. We simulated a deep water mixing event in a 20-L microcosm amended with 10% deep sea water (DSW), and monitored cell number, dynamics of community structure and DSW-responsive gene transcripts over the course of the 27-hr incubation.

Analysis of microbial transcript abundance over time suggested an immediate stimulation in gene expression of carbon fixation-related genes for *Prochlorococcus*. From T3 (12 hr) and onward, an *Alteromonas macleodii*-like population increased significantly in both cell abundance and the expression of genes involved in chemotaxis, cell motility, and C/N metabolism. In

contrast, the dominant heterotrophic bacterium *Pelagibacter* showed a relative decrease in relative transcript abundance, likely not due to transcriptional changes of *Pelagibacter* but instead due to the higher representation of fast-growing *Alteromonas*. This is consistent with the notion that *Pelagibacter* has a relatively small genome and a streamlined regulatory network (Giovannoni et al., 2005b) and so may be less responsive to fluctuations in ambient nutrient concentrations. Analyses also indicated intriguing phage dynamics in our data, pointing to potential presence of pseudolysogeny during deep mixing events. The prevalence of pseudolysogeny in nature remains to be elucidated, but its ecological implications are clear: nutrient loading in deep mixing events may affect host-phage interactions, triggering large-scale phage lysis that subsequently affects cell mortality and carbon cycling.

In summary, the experimental metatranscriptomic approach described here and in McCarren *et al* (McCarren et al., 2010) shows the potential for advancing our understanding of microbial processes and dynamics under specific environmental perturbations. We aimed to minimize artifacts in the microcosm experiments, by reducing incubation times, and introducing an unamended control as well as different types of treatments performed in parallel. Our findings set the stage for future inquiries on microbial community dynamics and metabolism associated with deep water mixing and the carbon cycles in the surface waters (McCarren et al., 2010). For example, based on the findings by McCarren *et al*, one can use *Alteromonas* and *Methylophaga* cultures to test the potential synergistic interactions between these species during HMWDOM degradation. Based on our observations of phage dynamics, testing the prevalence of pseudolysogeny using cyanobacteria model systems would be an interesting and worthy experiment (but potentially challenging due to the difficulty of mimicking nutrient-limiting conditions in the laboratory). In future experimental metatranscriptomic studies, the incorporation of more detailed chemical, physiological, and biochemical measurements (i.e., primary production, respiration rate, enzyme activity, nutrient concentration dynamics), will provide even more dimensions and resolution to data interpretations.

Tables and Figures

Table 1. Summary of database sizes, listed as the number of pyrosequencing reads. The removal of low-quality reads and identification of rRNA reads was described in Supplementary Methods. The exact reason for the consistently higher rRNA% in the Control cDNA samples was not clear, since the same rRNA subtraction protocol was used. Abbreviations: Con-Control; DSW-Deep Sea Water. T0: 0 hr; T1: 2 hr; T2: 6 hr; T3: 12 hr; T4: 19 hr; T5: 27 hr.

	Sample	Total # of reads (>50 nt)	Average read length (nt)	# of rRNA reads	rRNA%	Non rRNA reads	% of reads assigned to NCBI-nr	% of reads assigned to SEED
cDNA	Con T1	503302	194	395773	78.6	107529	34.7	25.2
	Con T2	476974	189	373455	78.3	103519	43.9	31.1
	Con T3	533875	198	438763	82.2	95112	51.4	36.1
	Con T4	596555	185	467921	78.4	128634	37.6	25.3
	Con T5	429400	185	332107	77.3	97293	44.2	30.2
	DSW T1	202745	178	116252	57.3	86399	56.4	44.7
	DSW T2	214438	184	141247	65.9	73191	53.0	40.5
	DSW T3	243398	181	155370	63.8	88028	53.9	39.7
	DSW T4	256186	181	159171	62.1	97015	59.2	43.3
	DSW T5	168419	188	109764	65.2	58655	60.4	44.5
DNA	Con T0	552689	245	2367	0.4	550322	66.7	43.0
	Con T5	418894	241	1599	0.4	417295	64.0	39.5
	DSW T5	519983	240	2179	0.4	517804	67.8	44.3

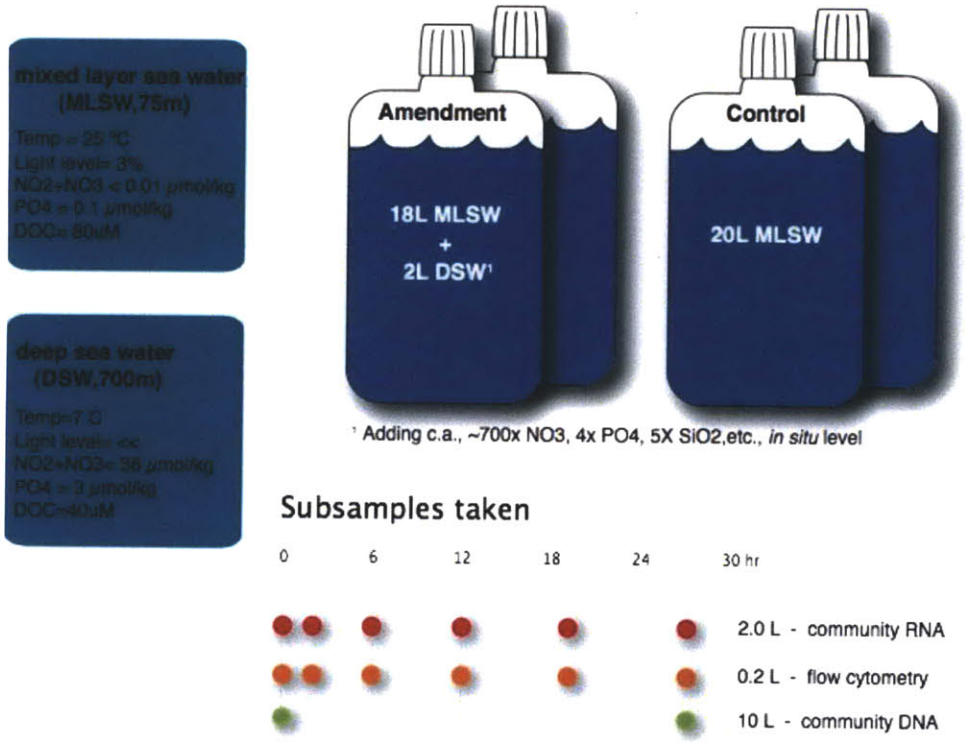


Figure 1. Deep sea water (DSW) amendment experimental setup and sampling regime. The experiment was performed in parallel with the DOM-amendment experiment previously reported (McCarren et al., 2010).

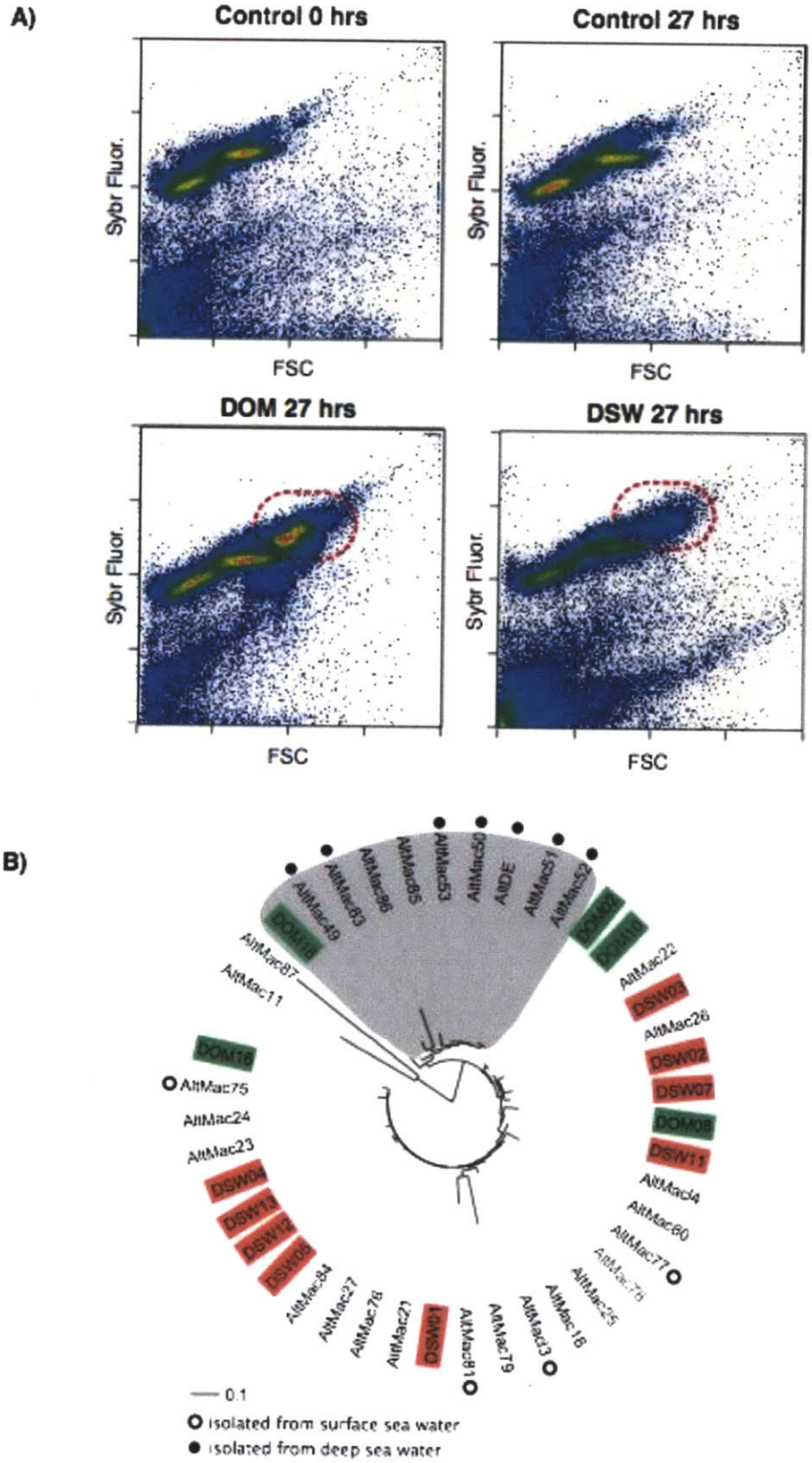

A)
Control 0 hrs
Control 27 hrs
DOM 27 hrs
DSW 27 hrs
Sybr Fluor.
FSC
B)
AltMac53
AltMac50
AltDE
AltMac51
AltMac52
AltMac86
AltMac83
AltMac49
DOM18
AltMac87
AltMac11
DOM02
DOM10
AltMac22
DSW03
AltMac26
DSW02
DSW07
DOM08
DSW11
AltMac4
AltMac60
AltMac77
AltMac75
AltMac24
AltMac23
DSW04
DSW13
DSW12
DSW05
AltMac84
AltMac27
AltMac78
AltMac21
DSW01
AltMac81
AltMac79
AltMac3
AltMac18
AltMac25
0.1
isolated from surface sea water
isolated from deep sea water

Figure 2 (Previous page). Flow cytometric and phylogenetic analysis of DSW-responsive heterotrophic bacterial populations. **(A).** Flow cytometry scatterplots for control and DSW treatment samples. DOM treatment sample is included for the purpose of comparison. The Control sample plot shows little changes in the distribution of cell size [as measured by forward scatter (FSC)] and DNA content (SYBR fluorescence) from beginning to end of the incubation. In contrast, most of the increase in cell numbers observed in the DSW-amended treatment can be attributed to the appearance of larger, high-DNA-content cells (circled in red). The same population (based on cell size and SYBR fluorescence) responded even more significantly to HMWDOM-amended treatment. **(B).** Phylogenetic reconstruction of near full length 16S rRNA sequences obtained from flow-sorted cells, together with those of *A. macleodii* isolates exported from SILVA SSU dataset (see Supplementary Methods). The cluster shaded in grey represents mostly deep ecotype *A. macleodii* strains, marked with solid black dots.

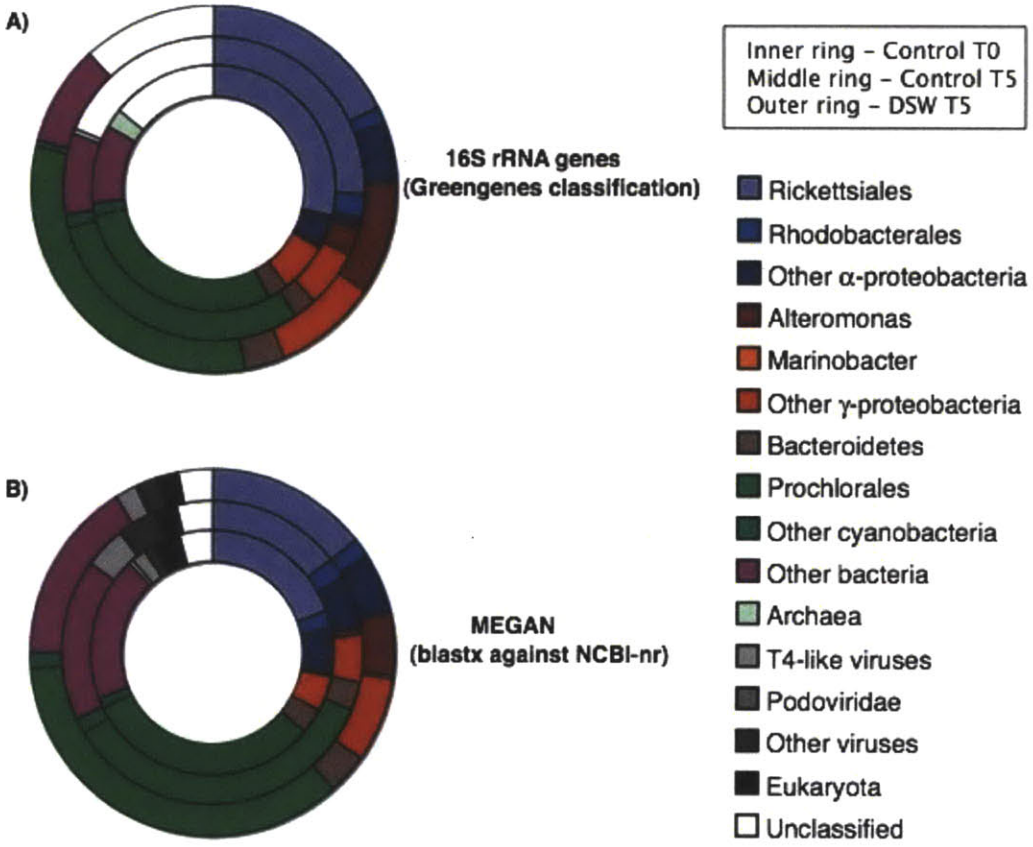


Figure 3. Microbial community composition assessed by taxonomic classification of 16S rRNA gene sequences and protein-coding mRNA sequences. Inner ring: Control initial time point DNA sample; middle ring: Control final time point DNA sample; Outer ring: DSW final time point DNA sample. **(A).** SSU rRNA reads classified by Greengenes taxonomy method (see Supplementary methods). **(B).** Protein-coding sequences classified using MEGAN (Huson et al., 2007). The percentages of mRNA reads that are presented here (with significant matches in NCBI-nr database) are listed in Table 1. Also note that MEGAN analysis revealed differential representation of phage-related sequences, which cannot be captured by 16S rRNA-centric analysis.

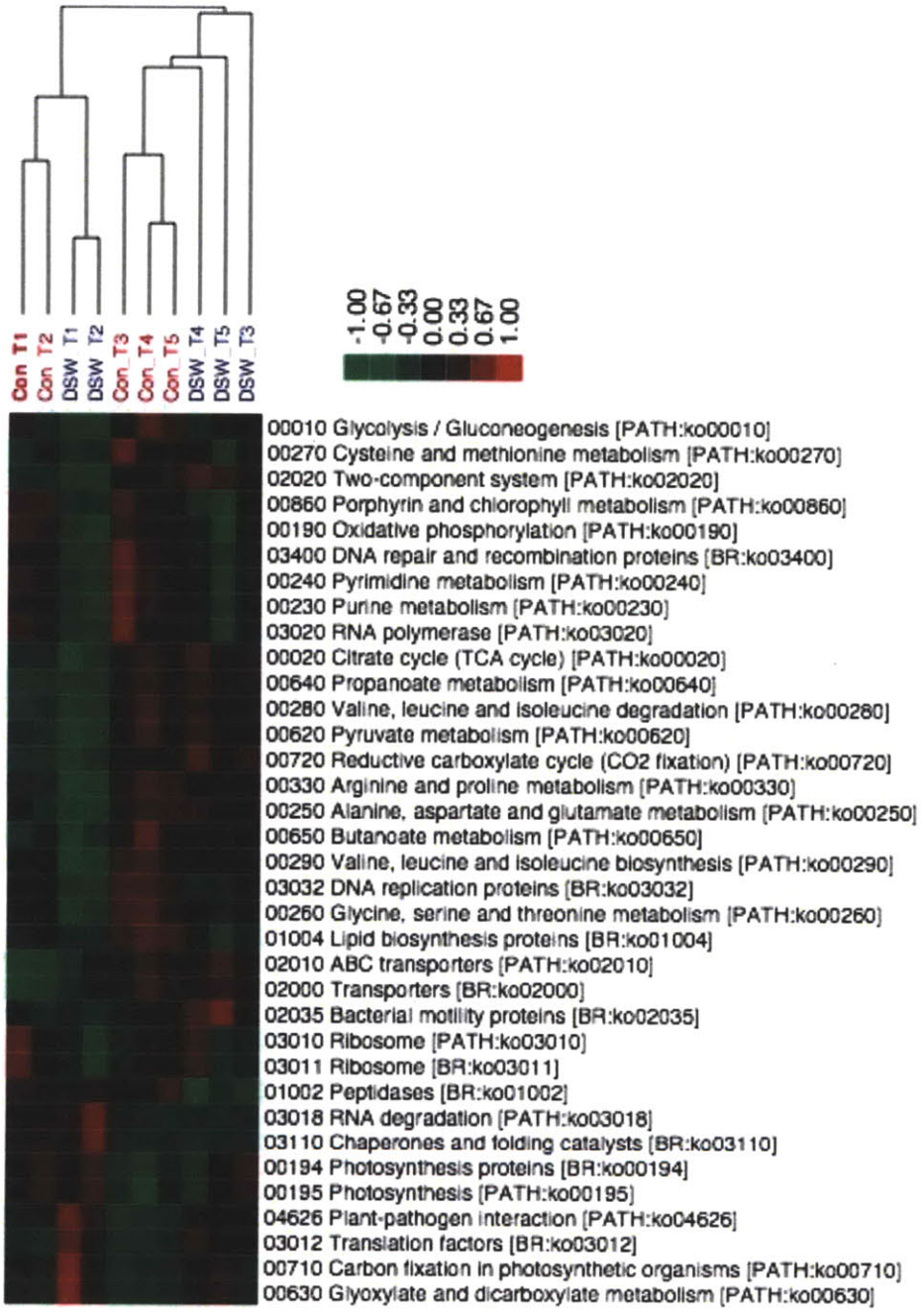


Figure 4. Clustering of 10 cDNA data sets based on relative representation of KEGG pathways (level 3 hierarchy). Dendrogram is based on hierarchical clustering of Pearson correlation coefficients for each pairwise dataset comparison, using the Genepattern workbench (Reich et al., 2006). The parameters used for clustering are: pathways that recruited ≥ 2% of total assigned reads at any one time point were used as input; data were centered and normalized for each pathway (mean = 0, squared sum = 1); hierarchical cluster with Pearson correlation (uncentered), and single linkage method.

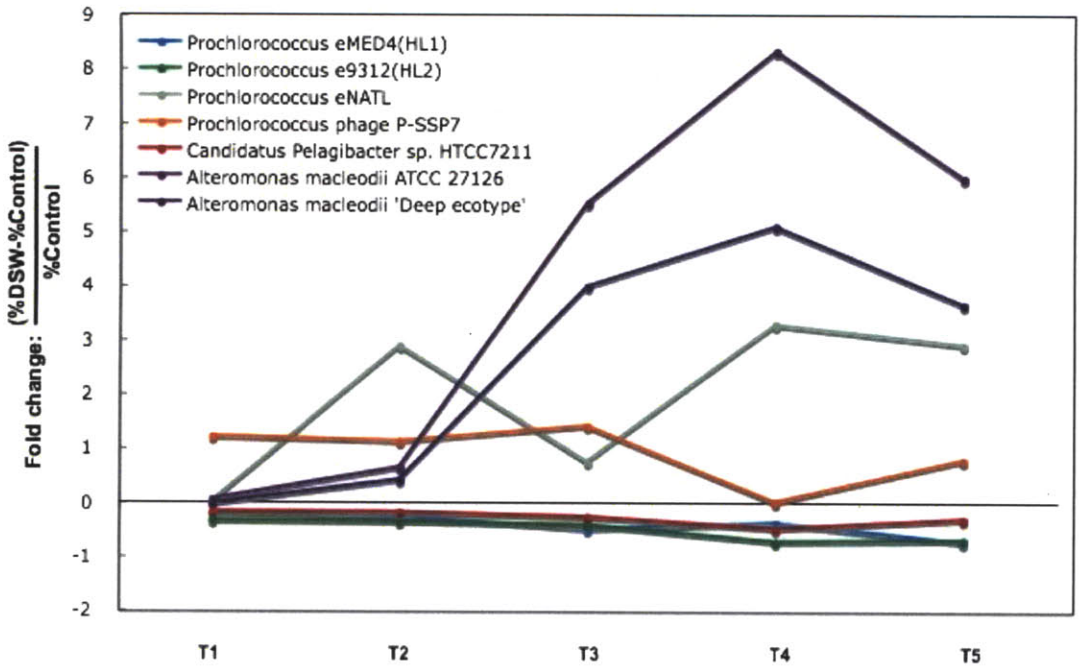


Figure 5. Comparison of genome-specific transcriptional activity between DSW amendment sample and the Control sample. A custom microbial genome database was used as reference, and cDNA reads were assigned to the top (or equally top) genome hit. For each genome, the relative representation (%) was defined as hit abundance normalized to the total number of reads assigned. Y-axis shows the normalized fold change of genome relative representation in the treatment relative to the control.

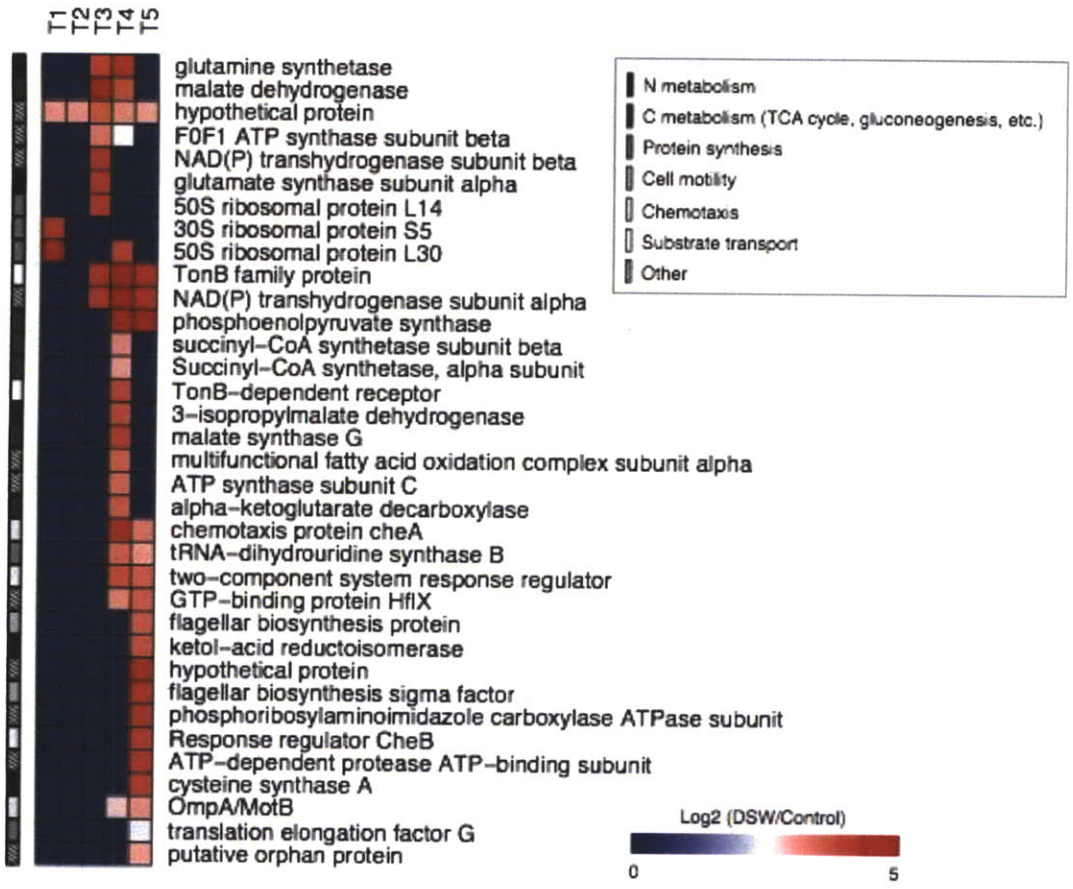


Figure 6. *Alteromonas* ORFs enriched in the DSW-amended sample, at least at one of the time points. ORFs were extracted from the *Alteromonas macleodii* ATCC 27126 genome. For each time point, differentially represented ORFs were identified using DEGseq at q-value ≤ 0.01 (see Supplementary Methods). Color on the plot indicates the level of enrichment in the treatment, blue to red being from lower to higher.

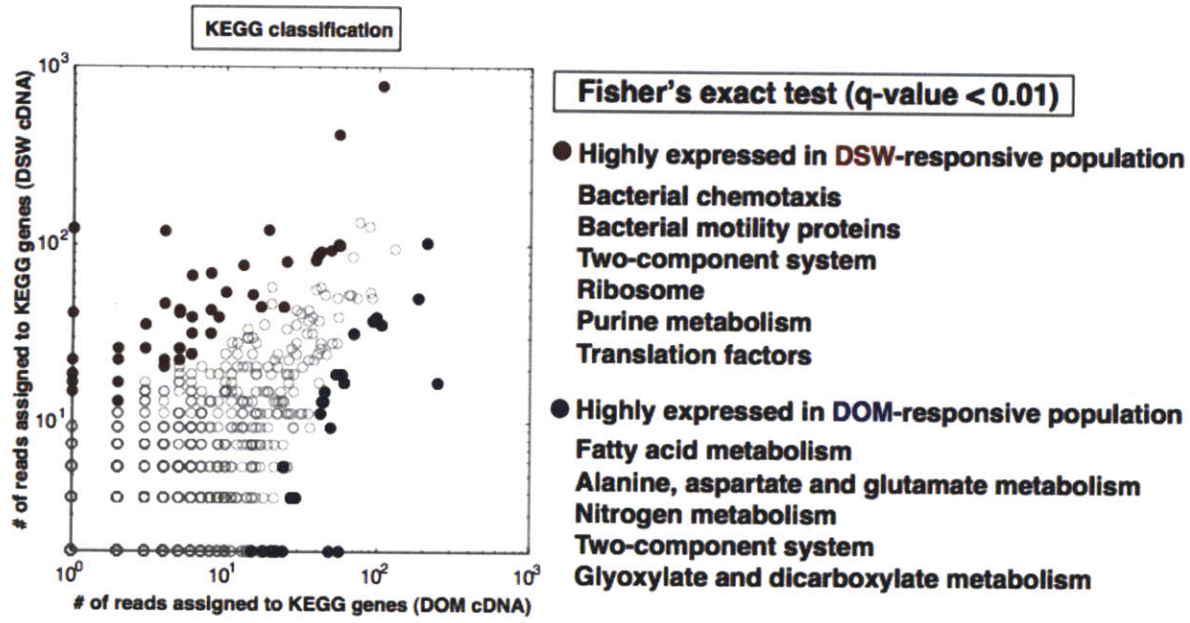


Figure 7. KEGG pathways that recruited significantly different number of *Alteromonas* cDNA reads in the DSW-responsive and DOM-responsive *Alteromonas* populations, in the T5 samples. Plotted here are KEGG reference genes with # of cDNA reads assigned. Fisher's exact test was used to identify KEGG genes with significantly different cDNA representation (q-value < 0.01;highlighted in red for DSW sample, and blue for DOM sample).

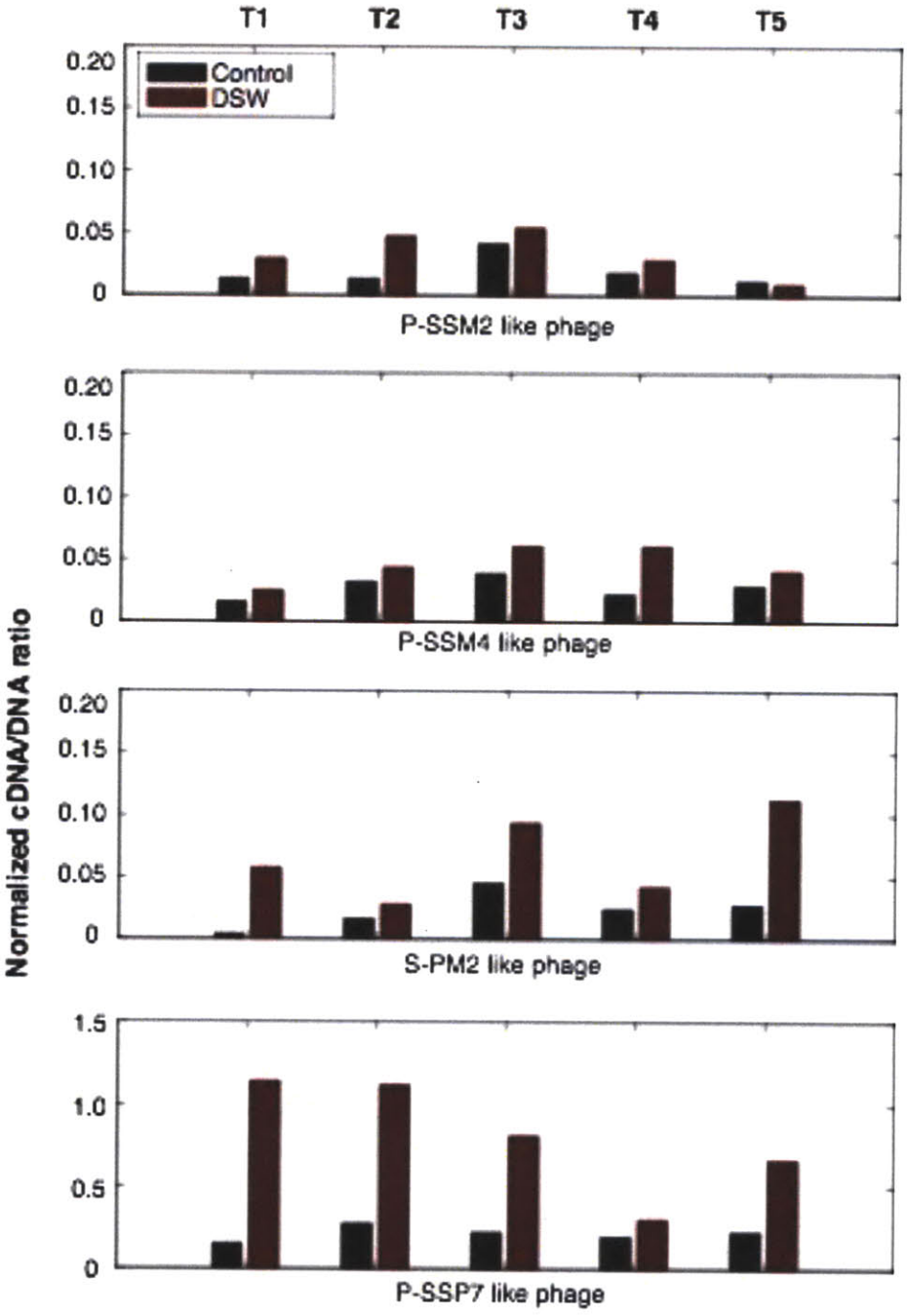


Figure 8. Normalized cDNA to DNA ratios for phage reference genomes, at each time point. Phage read sequences were identified using a more stringent set of criteria (see Supplementary Figure S7). Since only T0 and T5 DNA samples were sequenced, we used the average value of T0 and T5 phage DNA counts as the normalizer. Note that the scales of y-axes are different.

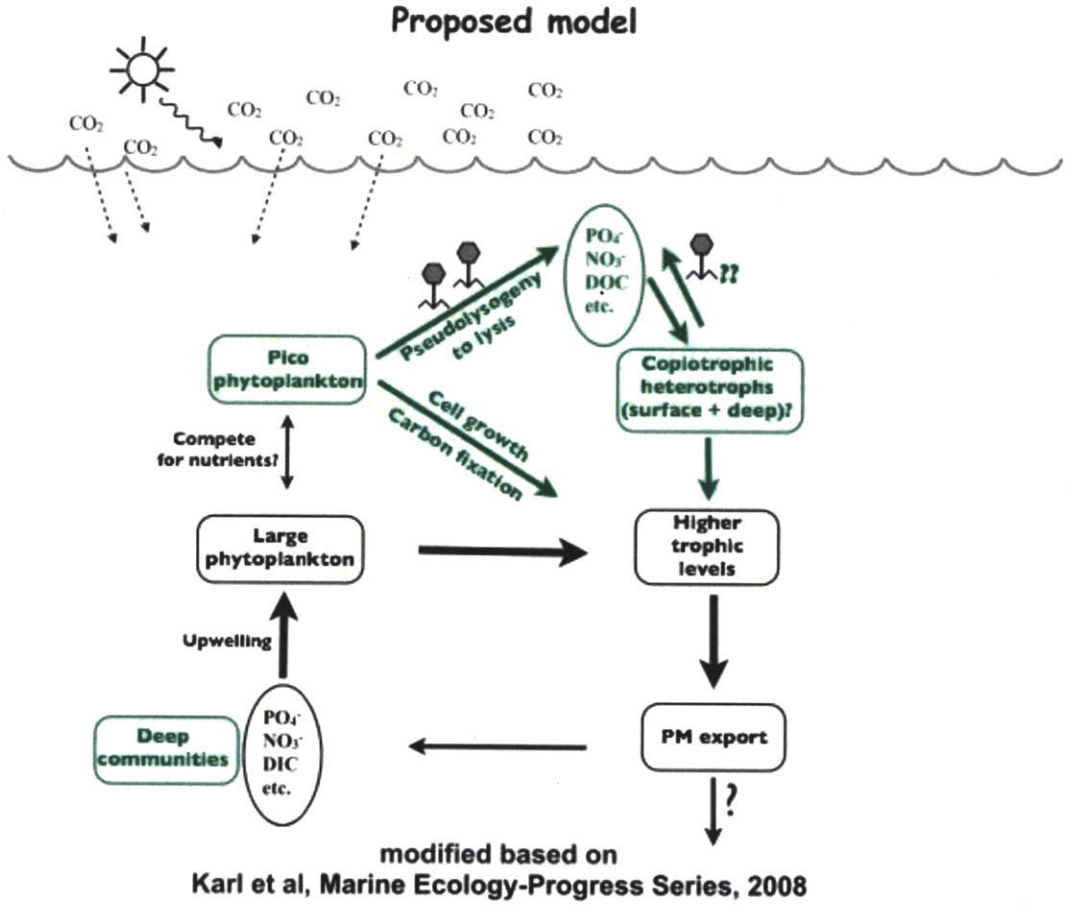


Figure 9. Proposed model of microbial responses in simulated deep mixing events. Model modified from Karl *et al* (Karl & Letelier, 2008).

Acknowledgements and author contributions

Experiments were performed, samples collected by Yanmei Shi, Jay McCarren, and Rex Malmstrom. We thank the captain and crew of the R/V Kilo Moana for facilitating sample collection, Chief Scientist Ricardo Letelier and all participants of the C-MORE BLOOMER cruise for help. We thank Rex for flow cytometry and rRNA sequencing. Many thanks to Rachel Barry for pyrosequencing library production and sequencing, and to John Eppley for computational assistance. This work was supported by the Gordon and BettyMoore Foundation (E.F.D.), the Office of Science–Biological and Environmental Research, US Department of Energy (E.F.D.), and National Science Foundation Science and Technology Center Award EF0424599.

Supplementary Information for Chapter 4

Supplementary Methods
Supplementary Tables S1-S3
Supplementary Figures S1-S8

Supplementary Methods

Sample collection and experimental setup

Seawater for on-deck microcosm incubation experiments was collected at 23°12.88′N, 159°8.17′W, from 75-m depth, pre-dawn, on August 16, 2007, during the CMORE BLOOMER cruise. Hydrocasts for sampling were conducted using a conductivity-temperature-depth (CTD) rosette sampler aboard the R/V Kilo Moana. Water was transferred to acid-washed, then sample-water rinsed 20L polycarbonate bottles. The deck-board incubator was a blue light type, which simulated the light levels at ~25-45m depth (roughly 14% surface irradiance). The carboys were wrapped in four layers of black fiberglass screen, to further decrease the light levels inside the carboy to ~3% surface irradiance, the *in situ* light intensity at 75m. These carboys were incubated in the deck-board incubators supplied with flow-through surface seawater to maintain near *in situ* temperatures (approximate 0.6° C temperature differential between 75m and sea surface over the course of experiment). In the same hydrocast, sea water from 700-m depth was collected, and brought up to the surface seawater temperature by immersing the bottle in the flow-through surface seawater for 10 min. For an initial total volume of 20 L, 2L of 700-m seawater was added to 18L of 75-m sample, roughly 700-fold increase of inorganic nitrogen (N) and 4-fold increase of inorganic phosphorus (P).

Replicate control and DSW amended microcosms were initiated at 05:45 local time with subsamples taken at 2, 6, 12, 19, and 27 hours post DSW addition. At selected time points, bacterioplankton biomass from ~2L seawater sample was rapidly collected for RNA samples by first pre-filtering through a 1.6μm glass fiber filter and then harvesting cells onto 0.2μm durapore (Millipore, Billerica MA). Filtration was limited to less than 10 minutes and then the filter was flash frozen in liquid nitrogen, immediately placed into *RNAlater* (Applied Biosystems, Foster City CA) and frozen at -80° C. Samples were transported frozen to the

laboratory in a dry shipper and stored at -80°C until RNA extraction procedures. RNA extraction, purification, and DNase treatments were performed as previously described (Frias-Lopez et al., 2008).

At both the beginning and the end of the experiment biomass in 10L water was similarly collected for DNA samples, first by pre-filtration through a 1.6 μm glass fiber filter and then collected onto 0.2μm Sterivex (Millipore) filters. Note that the 10L seawater for T0 DNA sample collection was directly taken from the CTD bottle, not from the microcosms. DNA extraction and purification performed as previously described (DeLong et al., 2006).

Flow Cytometry and Cell Sorting

At each time point 1 mL of seawater was preserved with 0.125% glutaraldehyde (final concentration), frozen in liquid N_2, and stored at –80°C for subsequent flow cytometric analysis and cell sorting using an Influx (Becton Dickinson). Prior to counting and sorting, samples were stained with SYBR Green (Invitrogen, Carslbad CA) for 15 min, and DNA-containing cells were identified based on fluorescence and scatter signals (Marie et al., 1997). Influx fluid lines were cleaned by running 10% bleach for 20 min followed by rinse with UV-treated MilliQ for 10min the previous night. Fluid lines where dried by pumping air through for 10 min before leaving overnight. Sheath fluid (1% NaCl w/v), sample tubes, and the sheath tank were UV-treated for 90min then left overnight, then re-treated with UV for 5 min the following morning.

A population of large non-pigmented cells appearing in DSW-amended incubations was sorted for identification by 16S rRNA gene sequencing. Approximately 7,000 cells from the final time point sample were first sorted into clean sheath fluid, and then re-sorted directly into 6 PCR tubes. In order to check contamination from the sheath fluid and samples lines, noise was sorted directly into a PCR strip tube, which were stored at –20 °C. Two rounds of sorting helped eliminate co-transport of dissolved DNA and ensured that only the targeted cells were amplified (Rodrigue et al., 2009).

Amplifications of 16S rRNA genes from flow-sorted cells were performed with universal 6F and 1492R primers, and the resulting amplification products pooled. These pooled PCR products were cloned using a TOPO-TA kit (Invitrogen, Carlsbad CA), and paired end reads sequenced using BigDye v3.1 chemistry on an ABI 3730 capillary sequencer (Applied

Biosystems, Foster City CA).

RNA Amplification, cDNA Synthesis, and pyrosequencing

Subtracted RNA was amplified using MessageAmp II (Ambion) following the manufacturer's instructions but substituting the T7-BpmI-$(dT)_{16}$VN oligo (GCCAGTGAATTGTAATACGACTCACTATAGGGGCGACTGGAGTTTTTTTTTTTTTTTT VN) in place of that supplied with the kit. The bases that are bold and underlined represent T7 promoter sequences.

Amplified RNA was then reverse transcribed into cDNA using SuperScript Double-Stranded cDNA Synthesis kit (Invitrogen) and random hexamer priming. Double-stranded cDNA was digested with BpmI to remove poly (A/T) tails. Before sequencing, poly (A/T)-removed cDNA was purified via the AMPure kit (Beckman Coulter Genomics, Danvers, MA, USA). Purified cDNA was used for the generation of single-stranded DNA libraries and emulsion PCR according to established protocols (454 Life Sciences, Roche). Clonally amplified library fragments were then sequenced on a Genome Sequencer FLX System (Roche).

Bioinformatics analysis

Removal of low quality reads and duplicate reads. A perl script was used to remove reads based on the report by Huse *et al* (Huse et al., 2007), that meet the criteria: 1) contain 3 or more "N"; 2) fall out of 95% distribution in length. Roughly 0.5% reads were removed using these criteria. The software cd-hit (Li & Godzik, 2006) was used to identify identical sequences in DNA samples. Roughly 3% of the remaining reads after quality control were identified as identical reads and removed. We did not remove identical reads from cDNA data sets, because it is impractical to assess if the duplicate sequences are artifacts or not.

Near full-length SSU rRNA gene amplicon sequences: Nine full length 16S rRNA gene sequences were obtained from flow sorted cells, and were aligned and classified using the Greengenes (DeSantis et al., 2006) NAST aligner and classification tool. Resulting alignments were compared with the SILVA (Pruesse et al., 2007) SSU rRNA database using ARB. For *Alteromonas*-specific phylogenetic analysis, full-length 16S rRNA sequences from a total of 29 *Alteromonas* isolates were exported from SILVA database. The weighted neighbor-joining tree was constructed using ARB, and viewed using tree-viewing tools on the Interactive Tree of Life

web site (Letunic & Bork, 2007).

Taxonomic analysis based on rRNA and protein-coding FLX reads: Identification of rRNA and protein-encoding reads were performed as described in Chapter 3, except that the BLASTx bits score cutoff used here was 50, due to longer read length. Greengenes (DeSantis et al., 2006) was used to align and classify 16S rRNA gene reads in the DNA samples; MEGAN (Huson et al., 2007) was used to extract taxonomic information from BLASTx output against NCBI-nr database (default parameters except minimum bits score of 50).

Functional gene analysis: Non-rRNA sequences were compared to NCBI-nr and KEGG databases using BLASTX for functional gene analyses. cDNA hit counts per NCBI-nr reference gene and per KEGG pathway (level 3 hierachy) were normalized to the total reads that matched the database used. NCBI-nr reference genes with significantly different counts between the treatment and control were identified using the R package DEGseq (Wang et al., 2010), under the following settings: FET (Fisher's Exact Test), q-value (a measure of significance in terms of false discovery rate) of 0.01. These differentially expressed nr reference genes were then classified to one of 12 most represented strains, based on NCBI taxonomy.

Relative representation of KEGG pathways was used to cluster 10 cDNA data sets using GenePattern workbench (Reich et al., 2006). Pathways that recruited $\geq 2\%$ of all assigned reads at any time point were used for hierarchical clustering using single linkage method, based on Pearson correlation coefficients for each pairwise dataset comparison, with data centered and normalized for each pathway (mean = 0, squared sum = 1).

Genome-centric analysis: A custom microbial genome database (ORF amino acid) was constructed from publicly available 2067 microbial genome sequences (as of January 2009), and was used to recruit cDNA reads. Reads with top hits with bits scores > 50 were assigned to the corresponding genomes. We then pooled all cDNA sequence reads assigned to a target genome, and compared the representation of each ORF on the genome in the treatment and control cDNA data. Differentially represented ORFs on the genome were identified for each time point data, using DEGseq as described above.

Comparison of DSW- and HMWDOM-responsive *Alteromonas* populations: Reads that were assigned as *Alteromonas*-related were defined as those with top BLASTx hit to

Alteromonas, with bits score ≥ 50. First, *Alteromonas* DNA reads were retrieved from DSW, DOM, and Control data sets, and compared against two *A. macleodii* reference genomes using BLASTn. The resulted BLASTn HSPs were used to calculate sequence identity distribution of the alignments. Next, we asked if we could detect gene content differences between DSW- and DOM-responsive *Alteromonas* populations. ORFs on the two *Alteromonas* reference genomes (AltDE and AltATCC) were categorized as shared (best reciprocal hits, with ≥ 50% aa identity, and ≥ 70% of the shorter read length), AltDE-specific, or AltATCC-specific. *Alteromonas* reads in the DSW and DOM T5 DNA data sets were assigned to these ORFs; based on the hit counts we identified differentially represented ORFs using DEGseq as described above. Finally, *Alteromonas*-related cDNA reads in DSW and DOM T5 cDNA data sets were assigned to KEGG pathways, and those with different relative representation was identified using DEGseq.

Supplementary Tables and Figures

Table S1. Nutrient concentration in the microcosm. Data were obtained from the BLOOMER website at: ftp://ftp.soest.hawaii.edu/dkarl/cmore/water/bloomer1/bloomer1.gof. Due to data limitation, nutrient concentrations at 700-m depth were sometimes extrapolated from data available for nearby depths.

	N (μM)	P (μM)	N: P	Si (μM)	DOC (μM)
75m water	0.01	0.1	0.1	2.5	77.3
700m water	38.3	2.8	13.7	61.7	44.3
Incubation initial condition	3.8	0.4	10.4	8.4	74.0
Factor increase	767.3	3.7	207.7	3.4	0.96

Table S2. Flow cytometric analysis of the Control and DSW amendment samples over time. Pro: Prochlorococcus. Total: total cell counts based on SYBR Green staining. Flow cytometry data were provided by Rex Malmstrom.

Sample	Cell type	Cell counts after accounting for dilution (cells/ml)						% change at 27 hr compared to 0 hr
		0 hr	2 hr	6 hr	12 hr	19 hr	27 hr	
Con	Pro	2.50E+05	2.41E+05	2.35E+05	2.29E+05	2.21E+05	2.25E+05	-10.0
Con	Total	7.56E+05	6.76E+05	6.92E+05	6.93E+05	7.07E+05	7.23E+05	-4.3
DSW	Pro	2.25E+05	2.24E+05	2.19E+05	2.23E+05	2.14E+05	2.15E+05	-4.6
DSW	Total	6.84E+05	6.34E+05	6.31E+05	6.56E+05	6.76E+05	7.54E+05	10.2

Table S3. # of reads that were assigned as *Alteromonas*, which were defined as reads with a top BLASTx hit against the NCBI-nr database to *Alteromonas*, with a bits score cutoff of 50.

Treatment	Data type	T5 (27 hr)
DSW	cDNA	8253
	DNA	28540
DOM	cDNA	11411
	DNA	44997
Con	cDNA	2204
	DNA	4706

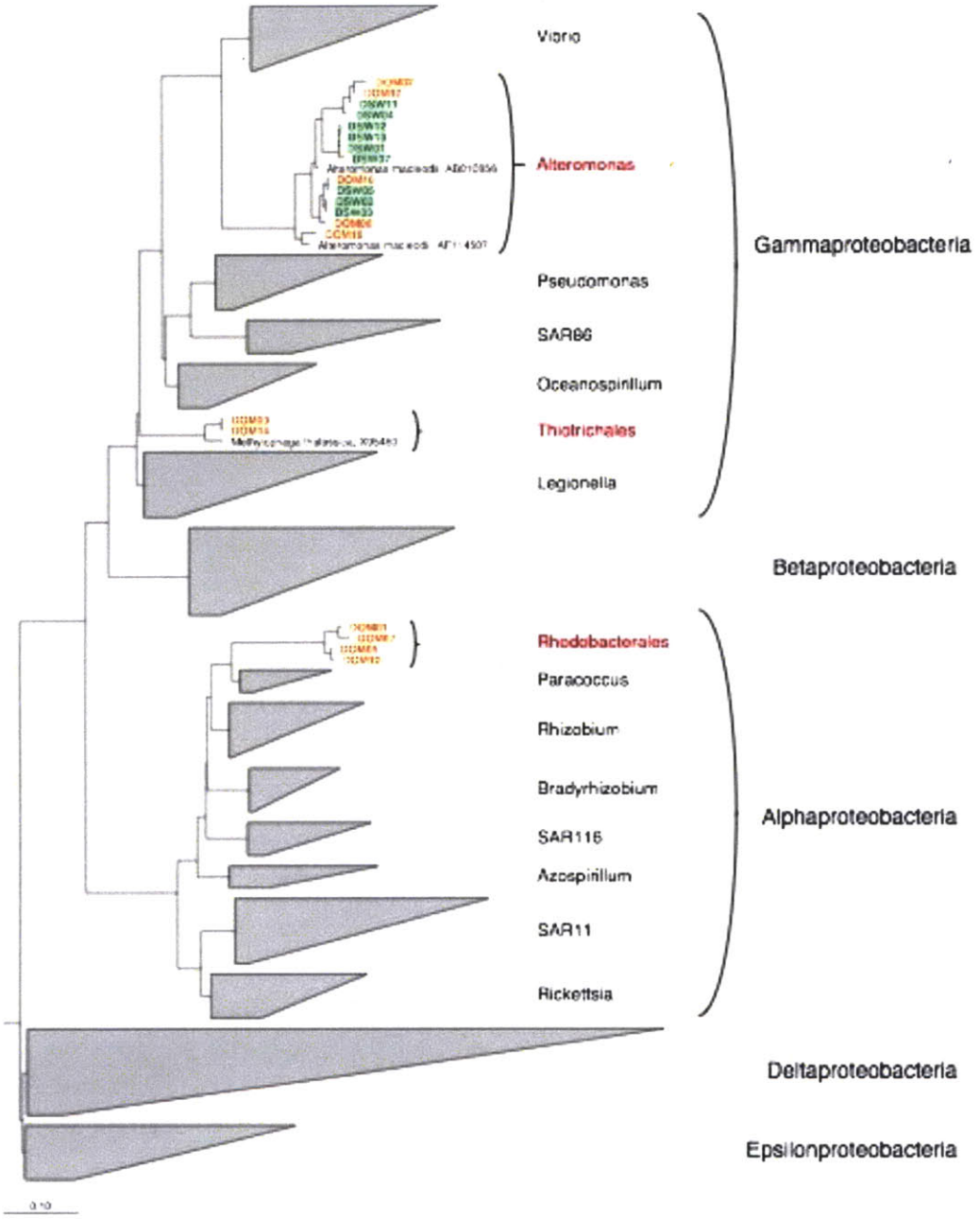


Figure S1. Phylogenetic tree (weighted Neighbor-joining) of selected SSU rRNA gene sequences from proteobacterial type strains, and the near full length SSU rRNA amplicon sequences obtained from flow cytometric sorting of the larger, higher-DNA-content population of cells present after DSW and HMODOM amendments.

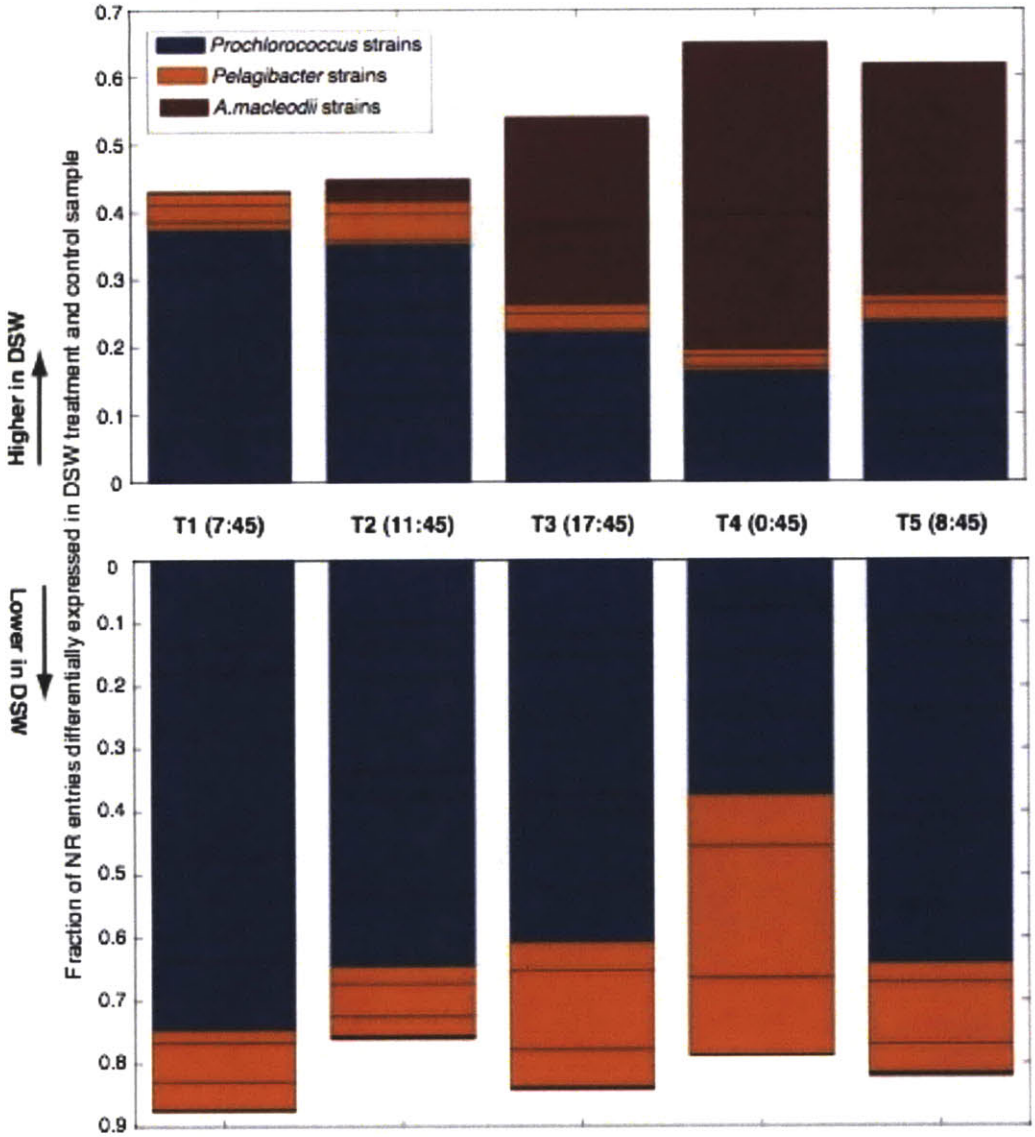


Figure S2. Putative taxonomic distribution of differentially represented NCBI-nr reference genes, in the cDNA datasets at each time point. cDNA reads were assigned to NCBI-nr reference genes using BLASTx, and hit counts were used to identify differentially represented nr reference genes using DEGseq (see Supplementary Methods). Identified nr reference genes were then assigned to a putative taxon based on NCBI taxonomy. Upper panel shows the taxa distribution of DSW-enriched nr reference genes, and lower panel DSW-depleted nr reference genes. Both y-axes represent the fraction of differentially represented nr reference genes assigned to a specific taxon out of the total identified. Only taxon with more than 20 differentially represented nr reference genes were plotted, including: *Prochlorococcus* strains MIT9202, MED4, MIT9312, AS9601, MIT9515, MIT9301, MIT9215; *Pelagibacter* strains HTCC1062, HTCC7211, and HTCC1002; *Alteromonas macleodii* strains "Deep ecotype", and ATCC27126.

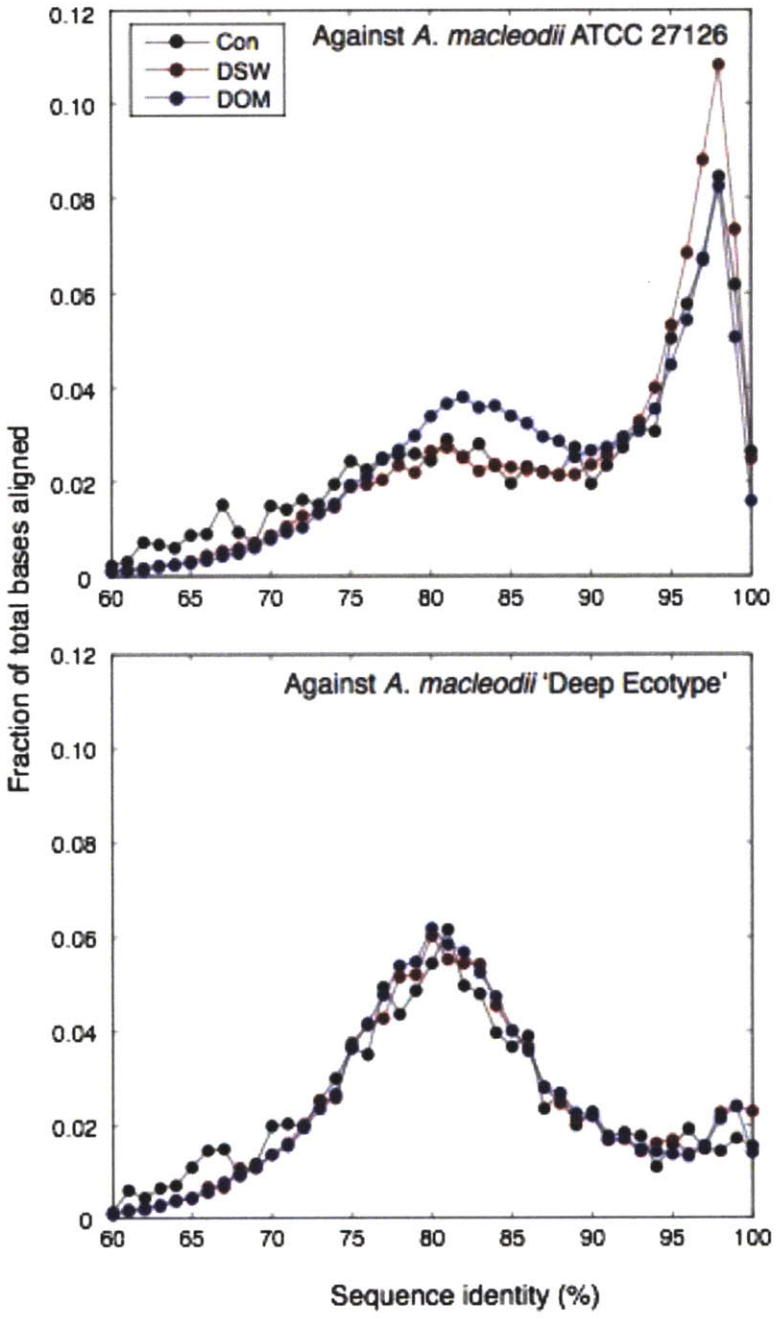


Figure S3. DNA sequence similarity of DSW-responsive, DOM-responsive, and Control *Alteromonas* populations, to the reference *A. macleodii* genomes. The plots indicate the fraction of total aligned base pairs to the reference genome by *Alteromonas* DNA reads (y axes) per unit of nt identity (x axes).

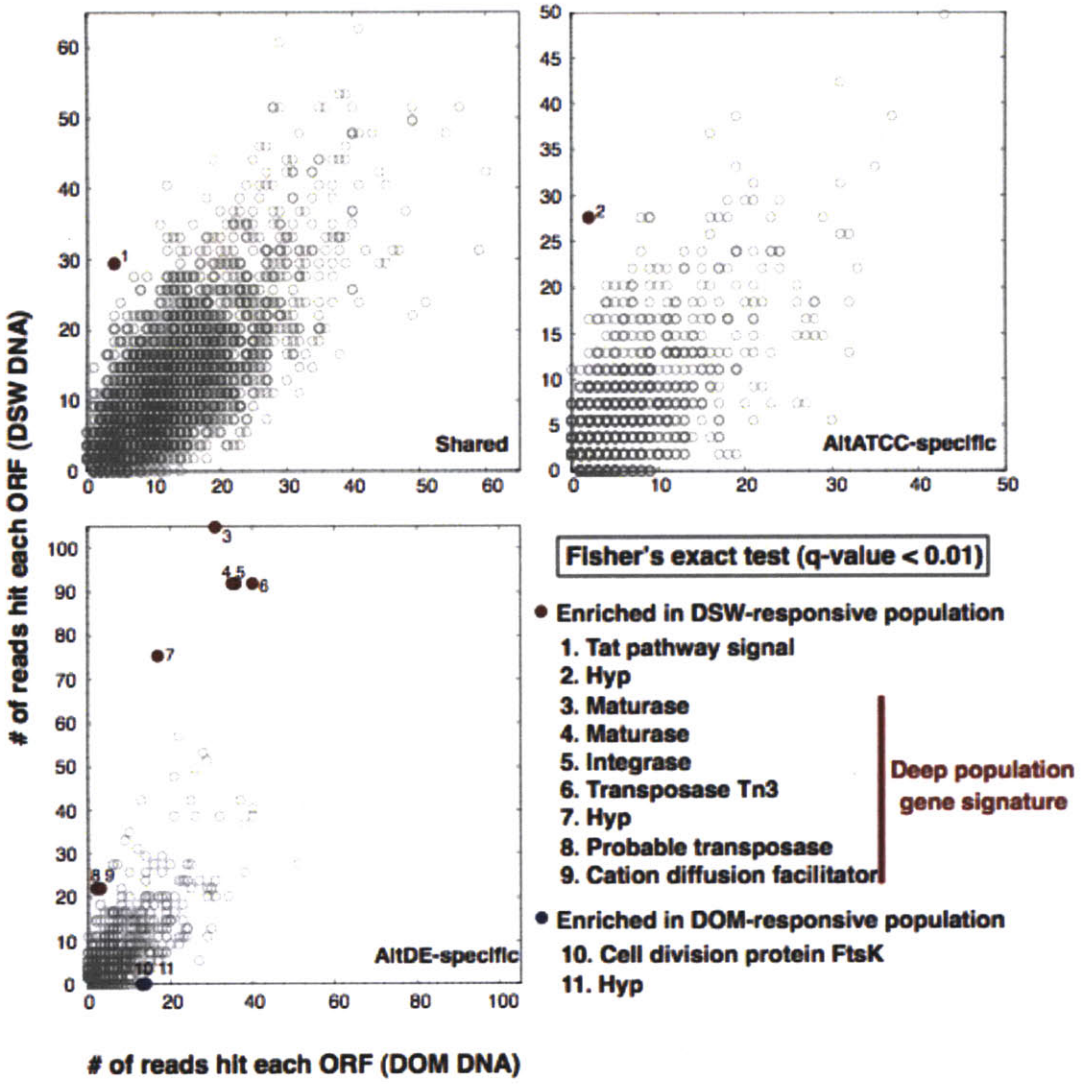


Figure S4. Detection of gene content differences in the DSW- and DOM-responsive *Alteromonas* populations. ORFs of the two *A. macleodii* genomes were divided into shared, AltDE-specific, and AltATCC-specific (see Supplementary Methods). ORFs with significantly difference abundance in *Alteromonas* T5 DNA data sets were highlighted: red for DSW sample, and blue for DOM sample

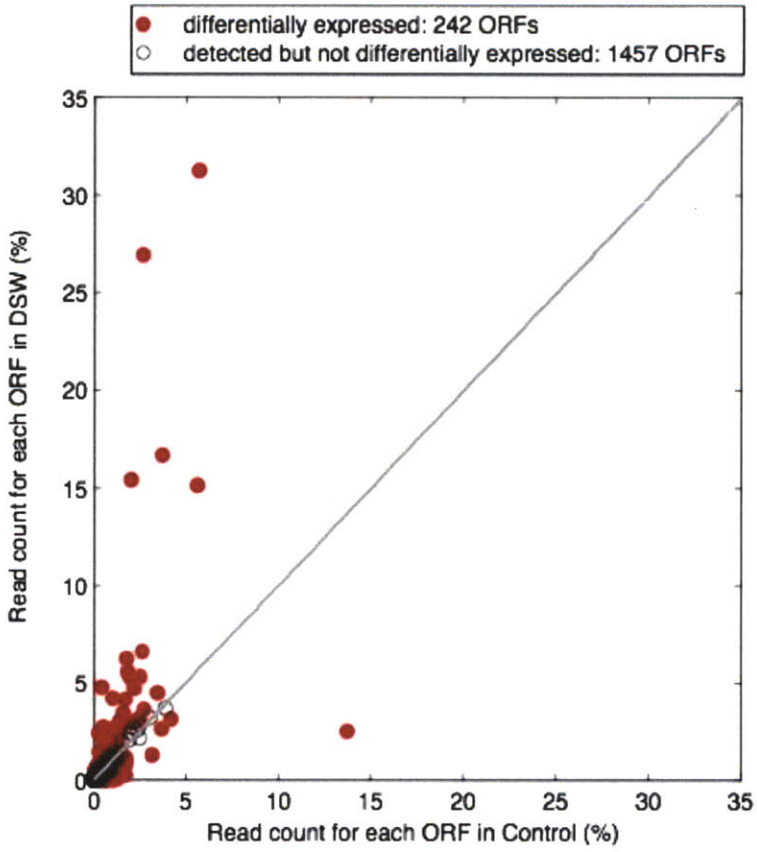


Figure S5. Illustration of ORF relative representation in DSW and Control cDNA samples. The *Prochlorococcus* strain AS9601was used as a reference in this analysis. ORFs with significantly different representation in the treatment and control were marked in solid red circles. ORFs detected in the data sets but not considered as differentially represented were marked in open black circles. DEGseq was used for evaluating statistical significance (see Supplementary Methods). Data for all time points were pooled in the figure.

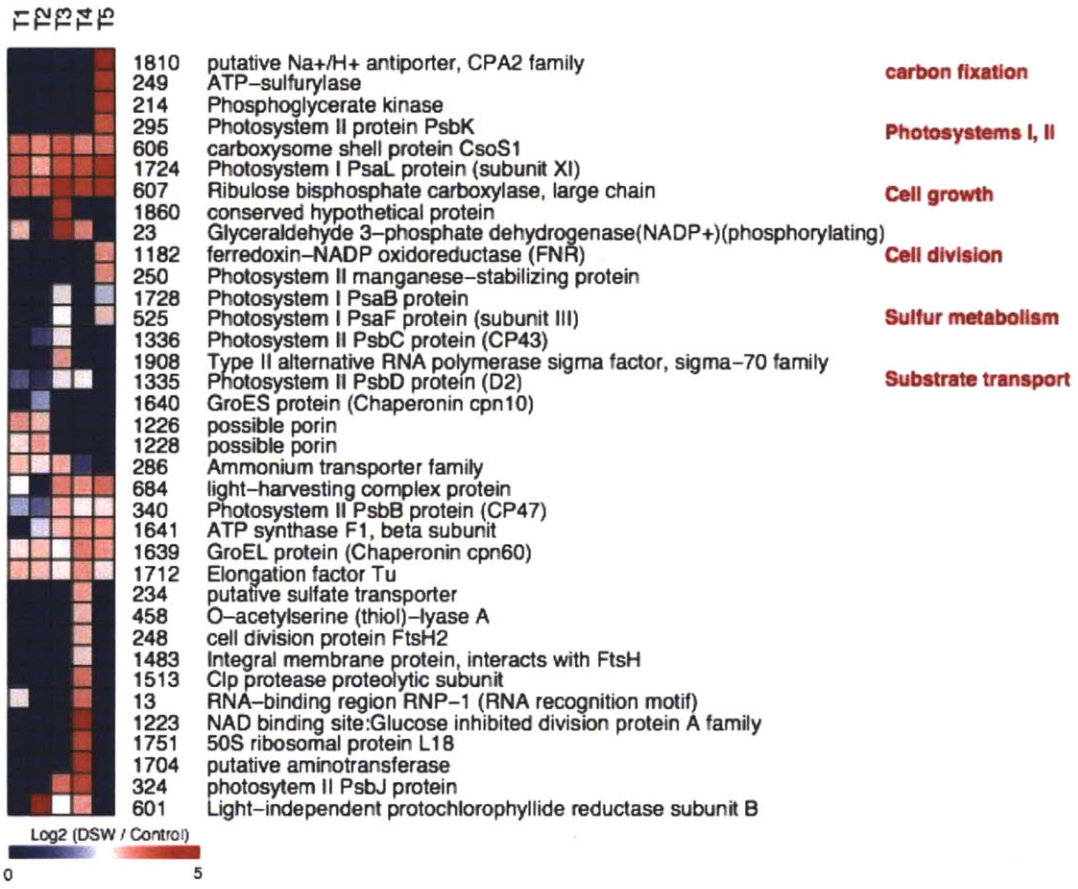


Figure S6. *Prochlorococcus* ORFs enriched in the DSW-amended sample, at least at one of the time points. ORFs were extracted from the *Prochlorococcus* AS9601 genome. For each time point, differentially represented ORFs were identified using DEGseq at q-value ≤ 0.01 (see Supplementary Methods). Color on the plot indicates the level of enrichment in the treatment, blue to red being from lower to higher.

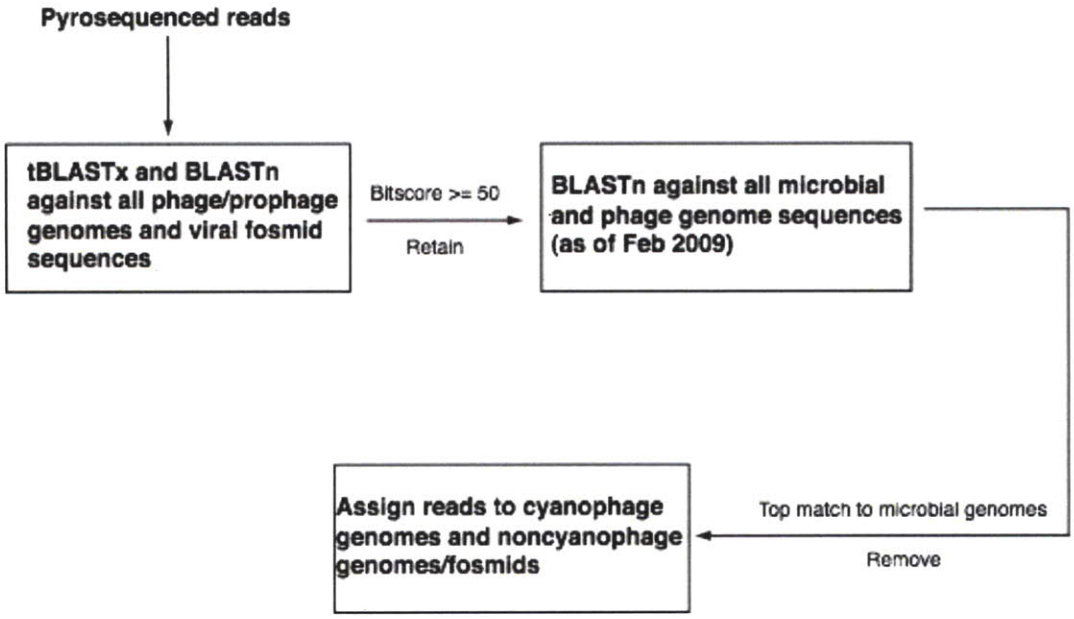


Figure S7. The flowgram showing criteria used for phage sequence identification. A more stringent set of criteria was used, because phage and host version of some protein-coding genes are indistinguishable at the amino acid level (Sullivan et al., 2006).

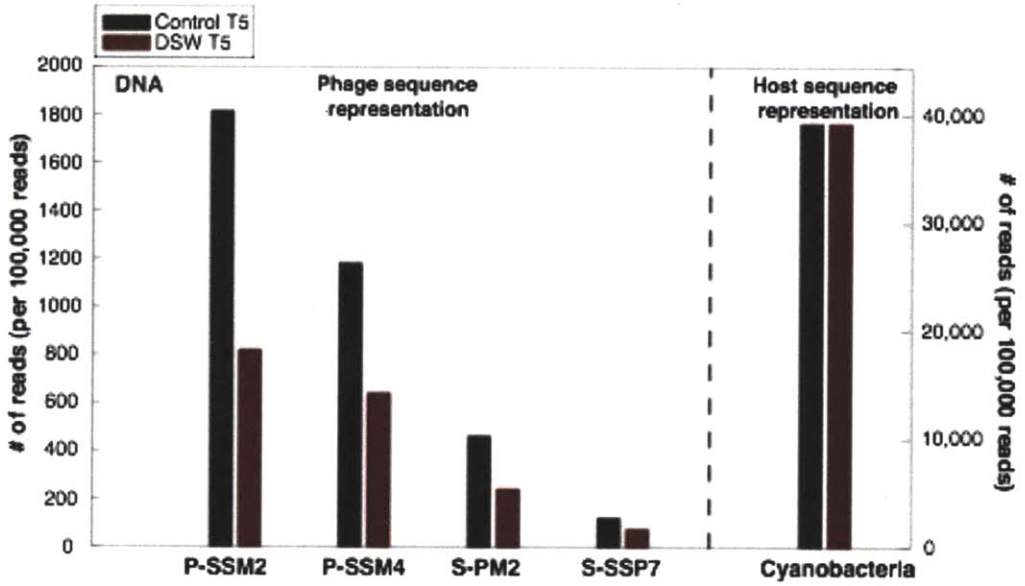


Figure S8. Representation of cyanophage like sequences in the Control T5 and DSW T5 DNA samples. Also presented (separated by the dashed vertical line) is the relative abundance of cyanobacteria like sequences. Note differences in the scales of the two y-axes.

CHAPTER FIVE

Metatranscriptomics reveals unique microbial small RNAs in the ocean's water column

Yanmei Shi, Gene W. Tyson, Edward F. DeLong

This chapter is presented, with slight formatting modification, as it appeared in *Nature* **459**:7244, 266-269 (2009). Corresponding supplementary information is appended.

Chapter 5: Metatranscriptomics reveals unique microbial small RNAs in the ocean's water column

Abstract

Microbial gene expression in the environment has recently been assessed via pyrosequencing of total RNA extracted directly from natural microbial assemblages. Several such 'metatranscriptomic' studies (Frias-Lopez et al., 2008; Gilbert et al., 2008) have reported that many cDNA sequences shared no significant homology with known peptide sequences, and so might represent transcripts from uncharacterized proteins. We report here that a large fraction of cDNA sequences detected in microbial metatranscriptomic datasets are comprised of well-known small RNAs (sRNAs) (Storz & Haas, 2007), as well as new groups of previously unrecognized putative sRNAs (psRNAs). These psRNAs mapped specifically to intergenic regions of microbial genomes recovered from similar habitats, displayed characteristic conserved secondary structures, and were frequently flanked by genes that suggested potential regulatory functions. Depth-dependent variation of psRNAs generally reflected known depth distributions of broad taxonomic groups (DeLong et al., 2006), but fine-scale differences in the psRNAs within closely related populations suggested potential roles in niche adaptation. Genome-specific mapping of a subset of psRNAs derived from predominant planktonic species like *Pelagibacter* revealed recently discovered as well as potentially new regulatory elements. Our analyses show that metatranscriptomic datasets can reveal new information about the diversity, taxonomic distribution and abundance of sRNAs in naturally occurring microbial communities, and suggest their involvement in environmentally relevant processes including carbon metabolism and nutrient acquisition.

Introduction

Microbial sRNAs are untranslated short transcripts that generally reside within intergenic regions (IGRs) on microbial genomes, typically ranging from 50-500 nucleotides in length (Storz & Haas, 2007). Most microbial sRNAs function as regulators, and many are known to regulate environmentally significant processes including amino acid and vitamin biosynthesis (Gottesman, 2002), quorum sensing (Lenz et al., 2004), and photosynthesis (Duehring et al., 2006). Since the identification and characterization of microbial regulatory sRNAs has relied primarily on a few model microorganisms (Silvaggi et al., 2006; Steglich et al., 2008; Vogel et al., 2003), relatively little is known about the broader diversity and ecological relevance of sRNAs in natural microbial communities.

During a microbial gene expression study comparing four metatranscriptomic datasets

from a microbial community depth profile (25m, 75m, 125m, 500m at Hawaii Ocean Time-series station ALOHA (Karl & Lukas, 1996)), we discovered that a significant fraction of cDNA sequences could not be assigned to protein-coding genes or ribosomal RNAs (rRNAs) (Figure 1). However, > 28% of these unassigned cDNA reads from each dataset mapped with high nucleotide identity (≥ 85%) to IGRs on the genomes of marine planktonic microorganisms (Supplementary Figure 1), suggesting they may be sRNAs. Consistent with the genomic location of known sRNAs (Kawano, Reynolds, Miranda-Rios & Storz, 2005), many of these reads mapped on IGRs distant from predicted ORFs, or were localized in clearly predicted 5'- and 3'-untranslated regions (UTRs).

Methods

Sample collection and RNA/DNA extraction

Bacterioplankton samples from the photic zone (25m, 75m, 125m) and the mesopelagic zone (500m) were collected from the Hawaii Ocean Time-series (HOT) Station ALOHA site in March 2006, as described previously (Frias-Lopez et al., 2008). Briefly, four replicate 1-liter seawater samples were prefiltered through 1.6-mm GF/A filters (Whatman, Maidstone, U.K.) and then filtered onto 0.22-μm Durapore filters (25mm diameter, Millipore, Bedford, MA) using a four-head peristaltic pump system. Each Durapore filter was immediately transferred to screw-cap tubes containing 1 ml of RNAlater (Ambion Inc., Austin, TX), and frozen at -80°C aboard the R/V Kilo Moana. Samples were transported frozen to the laboratory in a dry shipper and stored at -80°C until RNA extraction. Total sampling time, from arrival on deck to fixation in RNAlater was less than 20 minutes.

Total RNA was extracted as previously described (Frias-Lopez et al., 2008), using the *mir*Vana™ RNA isolation kit (Ambion, Austin, TX), with several modifications as follows. Samples were thawed on ice, and the 1 ml RNAlater was loaded onto two Microcon YM-50 columns (Millipore, Bedford, MA) to concentrate and desalt each sample. The resulting 50 l of RNAlater was added back to the sample tubes, and total RNA extraction was performed following the *mir*Vana™ manual. Genomic DNA was removed using a Turbo DNA-free™ kit (Ambion, Austin, TX). Finally, extracted RNA (DNase-treated) from four replicate filters were combined, purified, and concentrated by using the MinElute PCR Purification Kit (Qiagen,

Valencia, CA).

Bacterioplankton sampling for DNA extraction and DNA extraction was performed as previously described (Frias-Lopez et al., 2008).

Complementary DNA (cDNA) synthesis and sequencing

The synthesis of microbial community cDNA from small amounts of mixed-population microbial RNA was performed as previously described (Frias-Lopez et al., 2008). Briefly, nanogram quantities of total RNA were polyadenylated using *E. coli* Poly(A) Polymerase I (E-PAP) (Wendisch et al., 2001). First strand cDNA was then synthesized using ArrayScript™ (Ambion) with an oligo(dT) primer containing a T7 promoter sequence and a restriction enzyme (BpmI) recognition site sequence, followed by the second strand cDNA synthesis[1]. The double stranded cDNA templates were transcribed in vitro using T7 RNA polymerase at 37°C for 6 hours (Vangelder et al., 1990), yielding large amount of antisense RNA (aRNA). The SuperScript Double-Stranded cDNA Synthesis Kit (Invitrogen) was used to convert aRNA to microgram quantities of cDNA, which was then digested with BmpI to remove poly(A) tails. Purified cDNA was then directly sequenced by pyrosequencing (Margulies et al., 2005).

Removal of low-quality and ribosomal RNA (rRNA) GS20 cDNA sequences

Low quality cDNA reads were removed as previously described (Frias-Lopez et al., 2008). Reads encoding rRNA were identified and removed from the cDNA datasets by comparing them to a combined 5S, 16S, 18S, 23S, and 28S rRNA database derived from available microbial genomes and sequences from the ARB SILVA LSU and SSU databases (www.arb-silva.de). BLASTN (Altschul et al., 1990) matches with bit score $\geq$50 were considered significant and deemed rRNA sequences. In test simulations, this bit score cutoff resulted in <1.7% false positives against a database of all non-rRNA microbial genes from available microbial genomes.

Identification of protein-coding genes

Protein-coding cDNA reads were identified by translating nucleotide sequences in all 6 frames and comparing each to Global Ocean Sampling (GOS) peptides, the NCBI-nr protein

database, and a custom peptide database using BLASTX (Altschul et al., 1990). The custom peptide database contained marine specific open reading frame (ORF) sequences predicted from four sources: the Moore Microbial Genome Project genomes (http://www.moore.org/microgenome/strain-list.aspx), large genome fragments (~40 kb) from a variety of marine habitats (Rich *et al.*, in preparation), and both fosmid end sequences and shotgun library sequences generated from depth profile bacterioplankton samples collected in multiple HOT cruises (DeLong *et al.*, in preparation). Unpublished databases are available upon request.

After rRNA sequences were removed, each cDNA dataset contained between 40,000 to 70,000 pyrosequence reads. Of these cDNA reads, a large fraction (~50% of those from photic-zone samples; ~70% from the mesopelagic sample) showed no significant homology to either the non-redundant peptide database from NCBI or marine microbial peptide sequences, using the bit score of 40 that has been previously validated as a cutoff for calling homology in short pyrosequencing reads (Frias-Lopez et al., 2008).

Assignment of cDNA reads to known non-coding RNA families

We searched the Rfam database (Griffiths-Jones et al., 2005) to investigate the representation and diversity of known small RNA (sRNA) families in our datasets. Rfam is a collection of non-coding RNA families, represented by multiple sequence alignments and covariance models, including those from 400 complete genomes including 233 bacterial and 24 archaeal genomes (June 2008 version). The INFERNAL program (http://infernal.janelia.org/) was used to search for RNA structure and sequence similarities based on covariance models (CMs, also called profile stochastic context-free grammars) (Eddy & Durbin, 1994). The reference database was a collection of covariance models for all non-coding RNA families downloaded from the Rfam (version 8.1) ftp site (http://www.sanger.ac.uk/Software/Rfam/ftp.shtml). A perl wrapper named Rfamscan.pl (http://www.sanger.ac.uk/Software/Rfam/help/software.shtml), written by Sam Griffiths-Jones, was used to run batch queries (> 200,000 cDNA reads) on a local machine.

To test the specificity and sensitivity of the INFERNAL Rfam-seeded search of our cDNA reads, two datasets were created from the *Escherichia coli* strain K12 substrain MG1655,

in which sRNAs have been well defined (Rudd, 2000). The two test datasets were protein-coding sequences and known sRNA sequences, each with the same length distributions as our cDNA dataset (that is, 206,418 sequence fragments with mean sequence length 97bp). The INFERNAL Rfam-seeded search of the *E. coli* MG1655 protein-coding test dataset yielded no significant hits, suggesting high specificity and a false-positive rate below detection. However, the INFERNAL Rfam-seeded search did not identify all *E. coli* MG1655 sRNA fragments, likely due to the short lengths of the query sRNA fragments. To compensate for the decreased search sensitivity due to shorter read length, we queried all cDNA reads against all full length sRNA sequences in the Rfam database by BLASTN. Reads that did not meet the default cutoffs defined by Rfamscan, but shared good homology with Rfam member sequences by BLASTN (alignment length $\geq$ 90% of sequence length; sequence identity $\geq$ 85%) were also assigned to the corresponding sRNA families.

Putative taxonomic assignment of cDNA reads in known sRNA families

Potential taxonomic origins of the known sRNAs were investigated by searching against NCBI-nt (July 4th, 2008) using BLASTN (word size of 7, default e-value cutoff, low complexity filter off, and the ten best hits retained). The BLASTN results were then parsed using MEGAN (Huson et al., 2007) using default parameters, that is, the congruent taxonomy of the hits that were within 10% below the best hit was assigned to the cDNA read.

Self-clustering approach to identify sRNA and psRNA groups

A self-clustering approach allowed related cDNA reads to form distinct groups that could be separated from other transcripts based on sequence similarity and overall abundance. Combined cDNA reads (206,418 reads after the removal of rRNAs) from all four depths were locally aligned to each other (that is, all sequences served both as queries and subjects) using BLASTN with the following settings different from default: W = 7, F = F, m = 8, v = 206418, b = 206418, e = 1e-5. A perl script was used to group similar cDNA reads based on the BLASTN output. Briefly, for each cDNA query, all matches that met a minimum cutoff of 85% sequence identity over 90% average sequence length were considered significant and stored into a hash. The hash then was ranked based on the number of matches stored for each hash key (query). The cDNA read with the most matches served as a seed sequence of the first cluster. After all

matches of the seed sequence were recruited, the script looped over each one of the matches and gathered all subsequent matches until the chain disconnected and a new cluster started to form.

The self-clustering approach was successful in identifying a number of highly abundant psRNA groups. These psRNAs were clearly defined from protein-coding clusters as they were found in much higher copy number than most mRNAs, and the typical length of psRNAs was ~100-500 nucleotides. The sequence identity cutoff (85%) was chosen because it allowed known RNaseP RNAs from closely related microbial populations (for example, all *Prochlorococcus* RNaseP RNAs) to form a distinct sequence group. However, it is worth pointing out that since sRNA species by nature differ in their primary sequence divergence, clustering based on one sequence identity cutoff inevitably yields psRNA groups with different within-group diversity, which either represent homologs from closely related microbial populations or highly conserved elements from diverse microbial taxa.

Systematic screening for coding potentials of the self-clustered groups

We identified a total of 66 groups that contained more than 100 cDNA reads (a file named "H179_sRNA_groups.tgz", containing all sequences from these 66 groups, and a file named "H179_sRNA_groups_CLUSTAL.tgz", containing multiple sequence alignments of subsets of sequences from these 66 groups, can be downloaded from http://web.mit.edu/ymshi/Public/). To assess the possibility that some groups represent unannotated small proteins, we systematically screened multiple sequence alignments of these 66 groups for coding potentials based on 3-base periodicity in nucleotide substitution patterns. The rationale of detecting 3-base periodicity in coding regions is that codons encoding for the same amino acid often differ only in a single nucleotide located in the third position of the codon. As a direct consequence, in coding sequences under selective evolutionary pressure, substitutions are more often tolerated if they occur at the third position of codons. Therefore, if aligned sequences are protein-coding, the spectral signal of the mismatches along the alignment is expected to be maximal at frequency 1/3 (3-base periodicity) (Ré & Pavesi, 2007).

We generated a pipeline for multiple sequence alignment, nucleotide diversity calculation (conversion of DNA sequence alignments to numerical sequences), and Fourier Transform and power spectrum analysis of the numerical sequences, for all 66 groups (including known sRNAs

and psRNAs). Specifically, 100 sequences were randomly sampled from a subset of overlapping sequences in each group, and aligned using MUSCLE 3.6 (Edgar, 2004). The random sampling and alignment was repeated multiple times proportional to the number of sequences in the group. For each alignment, average nucleotide diversity was calculated for each column of the alignment as following:

$$D_{average} = \sum D_{pair-wise} / N(N-1)/2,$$

where $D_{average}$ represents average nucleotide diversity, $D_{pair\text{-}wise}$ represents pair-wise nucleotide diversity (a pair of identical nucleotides was given a value of 0, and a pair of different nucleotides was given a value of 1), and N(N-1)/2 represents the total number of pairs in the column of the alignment. Due to high insertion/deletion error rate of pyrosequencing (Margulies et al., 2005), any alignment column where greater than 75% of sequences had a gap resulted in that column being ignored in the subsequent calculation. After the multiple sequence alignments were converted to numerical sequences, a Fourier Transform and power spectrum analysis (Holste, Weiss, Grosse & Herzel, 2000) of the numerical sequences were performed using MATLAB (http://www.mathworks.com/) to find significant frequencies of periodicity.

Reverse transcription (RT)-qPCR analysis of psRNA Group 7 and sRNA Group 9

The apparent abundance and depth-dependant distribution of Group 7 and Group 9 in our metatranscriptomic datasets were validated using RT-qPCR. Due to lack of absolute quantification standards for these groups, we calculated their relative abundance to the crenarchaeal *amoA* transcript in the 500m sample. Primers for these groups were designed using the Invitrogen web-based OligoPefect primer designer. The primer sequences are: G7_Primer1 (AGCTCTGCTGGTTCYAGACT) and G7_Primer2 (TCGAACATTCACGCTTCCT); G9_Primer1 (TAAGCCGGGTTCTGTTCATC) and G9_Primer2 (GCCGCTTGAGACTGTGAAGT). The primer set for the crenarchaeal *amoA* transcript was the same as previously published (Mincer et al., 2007): CrenAmoAQ-F (5'-GCARGTMGGWAARTTCTAYAA), and CrenAmoAModR (5'-AAGCGGCCATCCATCTGTA). All primers were blasted against NCBI-nt database to avoid potential matches to unwanted regions.

Possible traces of DNA were removed from all RNA samples using Ambion's Turbo

DNA-free kit (Ambion, Austin, TX) following manufacturers instructions. For each reverse transcription (RT) reaction, 1 µl of RNA (4-7.5 ng) was reverse transcribed using gene-specific primer and Superscript III reverse transcriptase (Invitrogen, Carlsbad, CA). RT was performed at 50°C for 50 minutes, after an initial incubation step of 5 minutes at 65°C. The RT reactions were terminated at 85°C for 5 minutes, and 1µl RNase H was added to each RT reaction, followed by incubation at 37°C for 20 minutes. Subsequently, SYBR Green qPCR reactions were performed on LC480 (Roche Applied Science, Indianapolis, IN), using the specific primer set for each gene of interest. We used the $2^{-\Delta\Delta C_T}$ method (Livak & Schmittgen, 2001) to compare the relative abundance of Group 7 and Group 9 transcripts in all 4 samples (25m, 75m, 125m, and 500m) to the crenarcheal *amoA* transcript in the 500m sample.

Characterizing psRNA groups

The psRNA groups were further characterized to determine the approximate psRNA length, proximity to [5' or 3' or unknown (when the psRNA is not flanked by one ORF on each side)] and annotation of nearest flanking ORF on available genome/metagenome fragments, putative taxonomy and Support Vector Machine (SVM)-based RNA class probability. Pooled cDNA reads (not including rRNA reads) from each transcriptomic dataset were queried against a custom database of nucleotide sequences from available genome and metagenomic projects (see above) using BLASTN. Metagenomic fragments in this database were run through Metagene (Noguchi, Park & Takagi, 2006) to identify predicted open reading frames (coding) and intergenic (non-coding) regions.

Using the BLASTN and Metagene results, cDNA reads were mapped to each genome/metagenome fragment based on sequence similarity (≥85% identity over 90% of the read length), which could be used to calculate coverage values for each coding and intergenic region on each genomic/metagenomic fragment. Two groups were identified as highly expressed protein-coding genes (Group 35 - *amoC* and Group 42 - *amt*) and were excluded from further analyses. In most cases, reads belonging to putative sRNA groups mapped with high coverage to intergenic regions on genomic/metagenomic fragments. In these cases, we estimated the size of psRNAs in each group by defining the psRNAs as the sequence region in intergenic space having minimum sequence coverage of greater than 10X. In addition, it was also possible to determine the location of these psRNAs with respect to coding sequences. psRNAs were labeled

as either 3' or 5' based on their position relative to the nearest flanking gene. Functional annotation for each of the genes flanking psRNA groups was obtained by comparing the amino acid sequences against the KEGG (Kanehisa & Goto, 2000), COG (Tatusov, Galperin, Natale & Koonin, 2000) and the NCBI-nr databases from NCBI using BLASTP. Putative taxonomic origins of each fragment were assigned based on the NCBI taxonomy of matches in the NCBI-nr database.

Only 9 psRNA groups had no homology to sequences in currently available database. To estimate the size of each of these psRNA groups, reads from each were assembled using PHRAP (-minmatch 15, -minscore 20, revise_greedy) and the average length of contigs (<10 contigs) formed used to infer sequence space spanned by the sRNA group.

In order to calculate the RNA class probability for each group, the first twenty cDNA reads recruited to each psRNA group were extracted from the dataset and placed in the same sequence orientation. Multiple sequence alignments were performed using MUSCLE 3.6 (Edgar, 2004). The sequence alignment for each psRNA groups (CLUSTALW format) was then used to predict consensus structure and the thermodynamic stability using RNAz (Washietl, Hofacker & Stadler, 2005), and an RNA-class probability was calculated based on the SVM regression analysis.

Secondary structure prediction

The minimum free energy (MFE) structure was predicted based on the multiple sequence alignment of full-length psRNA sequences extracted from metagenomic sequence reads. The RNAalifold program from the Vienna RNA package (Hofacker, 2003; Hofacker, Fekete & Stadler, 2002) was used to produce consensus secondary structure and sequence alignment color-coded based on nucleotide variations. The color hue indicates how many of the six possible types of basepairs (GC, CG, AU, UA, GU, UG) occur in at least one of the sequences. Pairs without sequence covariation are shown in red. Ochre, green, turquoise, blue, and violet mark pairs that occur in two, three, four, five, and six types of pairs, respectively. Pale colors mark pairs that cannot be formed by all sequences (i.e., inconsistent base changes occur in some sequences). Attenuator-like structure was predicted using RibEx program (Abreu-Goodger & Merino, 2005).

Mapping cDNA reads to the genome of *Pelagibacter ubique* HTCC7211

Candidatus Pelagibacter ubique HTCC7211 genome sequences were downloaded from the Moore Microbial Genome Project (http://www.moore.org/microgenome/strain-list.aspx). Based on the genome annotations, all intergenic region (IGR) sequences greater than 50 bp (excluding rRNA and tRNA) were extracted and used to create BLASTN database. Both DNA and cDNA reads from each sample were then queried (BLASTN) against the database and parsed using same criteria as above (alignment length ≥90% of sequence length; identity ≥85%). For each IGR an expression ratio was calculated, as the percentage of cDNA reads assigned to the IGR, relative to that in the DNA library. If there were cDNA hits but no DNA hits, the number of DNA hits was considered as 1. This normalization compensates for the IGR length differences, and differences in DNA and cDNA library sizes.

Prediction of sRNA-containing IGRs in *Pelagibacter* genomes

Three *Pelagibacter* genomes (*Pelagibacter ubique* HTCC1062, HTCC1002 and HTCC7211) were used in the comparative genome analysis to predict possible sRNAs in the IGRs based on conserved secondary structure among closely related genomes (Axmann et al., 2005). A total of 1113 IGRs were extracted from above three genomes (again only IGRs ≥ 50bp and excluding tRNAs and rRNAs), and locally aligned to pooled ORFs and IGRs (5398) from the three genomes using BLASTN with the following settings changed from default: W = 7, F = F, v = 5398, b = 5398. ORFs were included so that cis-acting regulatory elements of mRNA were also examined. A total of 1848 IGR sequences were extracted from all the High-scoring Segment Pairs (HSPs) with bit scores greater than 50, using Bioperl (Jason & Ewan, 2000). Self-clustering of this subset of *Pelagibacter* IGR sequences was then performed, as described above. Sequences in each cluster were aligned using MUSCLE 3.6 (Edgar, 2004) and the alignments were scored for their secondary structure conservation and thermodynamic stability using RNAz 1.0 (Washietl et al., 2005). SVM-based RNA-class probability values from the RNAz pipeline were gathered for each cluster and ranked from high to low.

Results and Discussions

A covariance model-based algorithm (Eddy, 2007) was used to search all unassigned cDNA reads for both sequence and structural similarity to known sRNA families (Griffiths-Jones

et al., 2005). Thirteen known sRNA families were captured in the environmental transcriptomes, representing only ~16% of the total reads detected by IGR mapping. The most abundant sRNAs belonged to ubiquitous or highly conserved sRNA families including tmRNA, RNase P RNA, signal recognition particle RNA (SRP RNA), and 6S RNA (SsrS RNA) (Supplementary Table 1). In addition, a number of known riboswitches (cis-acting regulatory elements that regulate gene expression in response to ligand binding (Brantl, 2004)) were detected in lower abundance, including glycine, thiamine pyrophosphate (TPP), cobalamin, and S-adenosyl methionine (SAM) riboswitches (Supplementary Table 1). The apparent taxonomic origins of the most abundant known sRNAs revealed depth-specific variation that was generally, but not always, consistent with known microbial depth distributions (DeLong et al., 2006) (Supplementary Figure 2). For example, although SRP RNAs are abundant in our datasets, very few *Pelagibacter*-like SRP RNA reads were detected, suggesting that SRP-dependent protein recognition and transport may not be a dominant form of protein translocation in oceanic *Pelagibacter* populations.

To better characterize sRNAs in our datasets, including novel sRNA families (referred to as putative sRNAs (psRNAs) hereafter), we pooled all cDNA reads from each sample, and employed a self-clustering approach to group homologous cDNA reads (see Methods). Based on observations from the IGR mapping (Supplementary Figure 1), the self-clustering approach would help identify potential sRNAs since they are likely to span short genomic regions and exhibit high abundance (in many cases orders of magnitude higher than transcripts of protein-coding genes found in the same datasets). A total of 66 groups that comprised at least 100 overlapping cDNA reads were identified (Figure 2; Supplementary Table 2). For several of these groups, the abundance and depth-dependent distribution detected via cDNA pyrosequencing was confirmed using RT-qPCR analyses (Supplementary Figure 3). Among the 66 groups, 9 were identified as belonging to Rfam sRNA families (Supplementary Table 2), and the majority of the remaining psRNA groups mapped to IGRs on metagenomic fragments derived from marine planktonic microorganisms.

Although they bear no resemblance to known peptide sequences, the psRNA groups could potentially represent mRNA degradation products or small unannotated protein-coding regions. We applied several criteria to help rule out these possibilities, including location within IGRs, psRNA length, lack of coding potential, and conserved secondary structure. First, the

psRNAs ranged in size between 100 and 500 nucleotides (Supplementary Figure 4; Supplementary Table 2), and tended to have an elevated GC content when located within an AT-rich genome context (Schattner, 2002) (Figure 3A). Second, we systematically screened multiple sequence alignments of all 66 groups for coding potential, as indicated by 3-base periodicity in the nucleotide substitution patterns (Ré & Pavesi, 2007) (Methods). Only Group 92 was identified as possibly protein encoding (Figure 3B), and this was subsequently mapped to a hypothetical protein (ABZ07689) from a recently described uncultured marine crenarchaeote (Konstantinidis & DeLong, 2008). Third, the psRNA groups encompassed relatively divergent sequences that shared conserved secondary structures (e.g., Figure 3A inset), suggesting evolutionary coherence of functional roles and mechanisms. The alignment of full-length psRNA sequences revealed clear nucleotide co-variation that preserved base-pairing in the consensus secondary structure (e.g., Supplementary Figure 5). In a specific example (Group 5), while three divergent *Pelagibacter*-like psRNA sequences (one from 4000 m depth (Konstantinidis & DeLong, 2008) and two from surface waters (Rusch et al., 2007)) shared pairwise nucleotide identities of only 78% to 87%, yet predicted secondary structures were nearly identical (Supplementary Figure 6). Although computational analyses alone cannot be completely definitive, these combined criteria support our hypothesis that most psRNA groups we identified represent authentic microbial sRNAs.

Many of the psRNAs identified here may be derived from as-yet-uncharacterized microorganisms. For instance, nine self-clustered psRNA groups shared no obvious homology with known nucleotide sequences (e.g., Group 6 and 10), and appear to represent completely novel sRNA families. The majority of these were found only in the 500 m sample (Figure 2). The remaining psRNA groups mapped to IGRs on genomic and metagenomic sequences derived from planktonic marine microbes. Although identifying sRNA regulatory functions and their target genes is a major challenge even for model microorganisms (Vogel & Wagner, 2007), the conserved genomic context of these psRNAs has potential to provide insight into their functional roles (Hershberg, Altuvia & Margalit, 2003; Yao et al., 2007). The most predominant gene families flanking these psRNA groups included transporter genes involved in nutrient acquisition (inorganic nitrogen, amino acids, iron and carbohydrates), and genes involved in energy production and conversion (Supplementary Table 2). These results highlight the potential importance of sRNA regulation of nutrient acquisition and energy metabolism in free-living

planktonic microbial communities.

The most populated psRNA cluster, Group 4, appeared to be involved in the regulation of central carbon metabolism and energy production in *Proteobacteria* (predominantly *Gammaproteobacteria*). The psRNAs from this group were flanked by genes involved in pyruvate metabolism (e.g., pyruvate kinase and malate synthase), glucose transport (e.g., sodium glucose symporter), and nitrogen acquisition (e.g., ammonia permease and aminopeptidase) (Figure 2; Supplementary Table 2). In several cases, Group 4 psRNAs occurred in tandem copies within the same IGR (Figure 3A). Small RNAs that display stable secondary structure typically mediate regulation using sequences in loop domains to interact with specific target sequences (Storz & Haas, 2007; Trotochaud & Wassarman, 2005). Consistent with this mechanism, a conserved 6-nt sequence motif (AAGAGN) appeared in multiple loops within predicted hairpin structures for Group 4 (Figure 3A inset). The 6-nt sequence AAGAGA was previously verified as a ribosomal binding site (Bruttin & Brüssow, 1996), and suggests that Group 4 psRNAs may play a regulatory role at the translational level. Indeed, sequences in one of the loop domains of the consensus structure (Figure 3A inset) have potential to interact (by base pairing across 32 bps) with the flanking pyruvate kinase gene near the 5' translation initiation site.

In contrast to the broad taxonomic affiliations of Group 4 psRNAs, the other highly abundant psRNA group, Group 5, appeared almost exclusively on *Pelagibacter*-like genomic fragments recovered from both open ocean surface waters (Rusch et al., 2007) and abyssal (4000m) depth (Konstantinidis & DeLong, 2008), but did not map to the genomes of currently cultivated *Pelagibacter* strains (Figure 2; Supplementary Table 2). Group 5 psRNAs mapped onto 203 different metagenomic fragments, predominantly in the 5'-UTR of 6-O-methylguanine DNA methyltransferase (6-O-MGMT; COG0350; involved in DNA repair), and the 3'-UTR of tRNA (5-methylaminomethyl-2-thiouridylate)-methyltransferase (*trmU*; COG0482; involved in tRNA modification). A predicted promoter and Rho-independent terminator flanked Group 5 psRNAs upstream of 6-O-MGMT, and attenuator/riboswitch characteristics were identifiable in the 5'-UTR by secondary structure prediction (Supplementary Figure 6). Indeed, the presence of riboswitch-like elements upstream of 6-O-MGMT genes was previously predicted by comparing 223 complete bacterial genomes (Abreu-Goodger & Merino, 2005).

Unlike Group 4 and 5 psRNAs, the remaining self-clustered sRNA and psRNA groups

showed depth-variable distributions (Figure 2). Group 7 psRNAs were enriched at 500m and were highly conserved in marine crenarchaeal genomes. Similarly, *Cyanobacteria*-like psRNAs were enriched in the photic zone (e.g. Group 2, 30, 48 and 17; Supplementary Table 2). One of these groups (Group 30) includes two experimentally validated sRNAs (Yfr8 and Yfr9), which were found antisense to one another and were hypothesized to be involved in a toxin-antitoxin system in *Prochlorococcus marinus* MED4 (Steglich et al., 2008). Intriguingly, a few *Prochlorococcus*-like psRNA groups mapped to some but not all coexisting members of the *Prochlorococcus* population, suggesting that such sRNAs may provide niche-specific regulation. Group 2 psRNAs, for example, were detected only in the genome of *P. marinus* strain MIT9215, and in a highly similar genomic fragment from the environment (DQ366713). Group 2 psRNAs are located in a hyper-variable region adjacent to phosphate transporter genes, and share a 14-bp exact match with the 5' translation initiation site of the phosphate ABC transporter gene (*pstC*). In *Prochlorococcus* strains lacking the phoBR two-component regulatory system {Martiny 2006}, such as MIT9215, it is possible that sRNAs represent an alternative mechanism for regulating phosphorus assimilation.

To examine sRNA representation in specific abundant microbial groups, we aligned the psRNA reads to the genome of an abundant planktonic bacterium, *Candidatus Pelagibacter ubique* HTCC7211. Eleven IGRs on the *P. ubique* HTCC7211 genome coincided with the psRNAs identified in our samples (Figure 4), 6 of which were also independently predicted as sRNA-containing IGRs (SVM RNA-class probability > 0.9) by comparative analysis of three *P. ubique* genomes (Methods; Supplementary Table 3). Genes flanking these expressed psRNAs included DNA-directed DNA polymerase *gamma/tau* subunit (*dnaX*), *carD*-like transcriptional regulator family, and alternative thymidylate synthase (Supplementary Table 3). Notably, covariance model-based searches identified cDNAs mapping to glycine riboswitch motifs in two *Pelagibacter* IGRs (Figure 4; Supplementary Table 3). Recently, it was experimentally verified that *P. ubique* HTCC1062 uses one of these two glycine riboswitches to sense intracellular glycine level and to regulate its carbon usage for biosynthesis and energy (Tripp et al., 2008).

The diversity and abundance of sRNAs in microbial metatranscriptomic datasets indicates that natural microbial assemblages employ a wide variety of sRNAs for regulating gene expression in response to variable environmental conditions. The data and analyses described

here provide a culture-independent tool to expand our knowledge of microbial sRNA sequence motifs, structural diversity, and genomic distributions. Although the exact regulatory functions of many of the psRNAs remain to be experimentally verified, their *in situ* expression, their structural features, and their genomic context, all provide a solid foundation for future studies. These data, in conjunction with metatranscriptomic field experiments linking environmental variation with changes in RNA pools, have potential to provide new insights into environmental sensing and response in natural microbial communities.

Figures

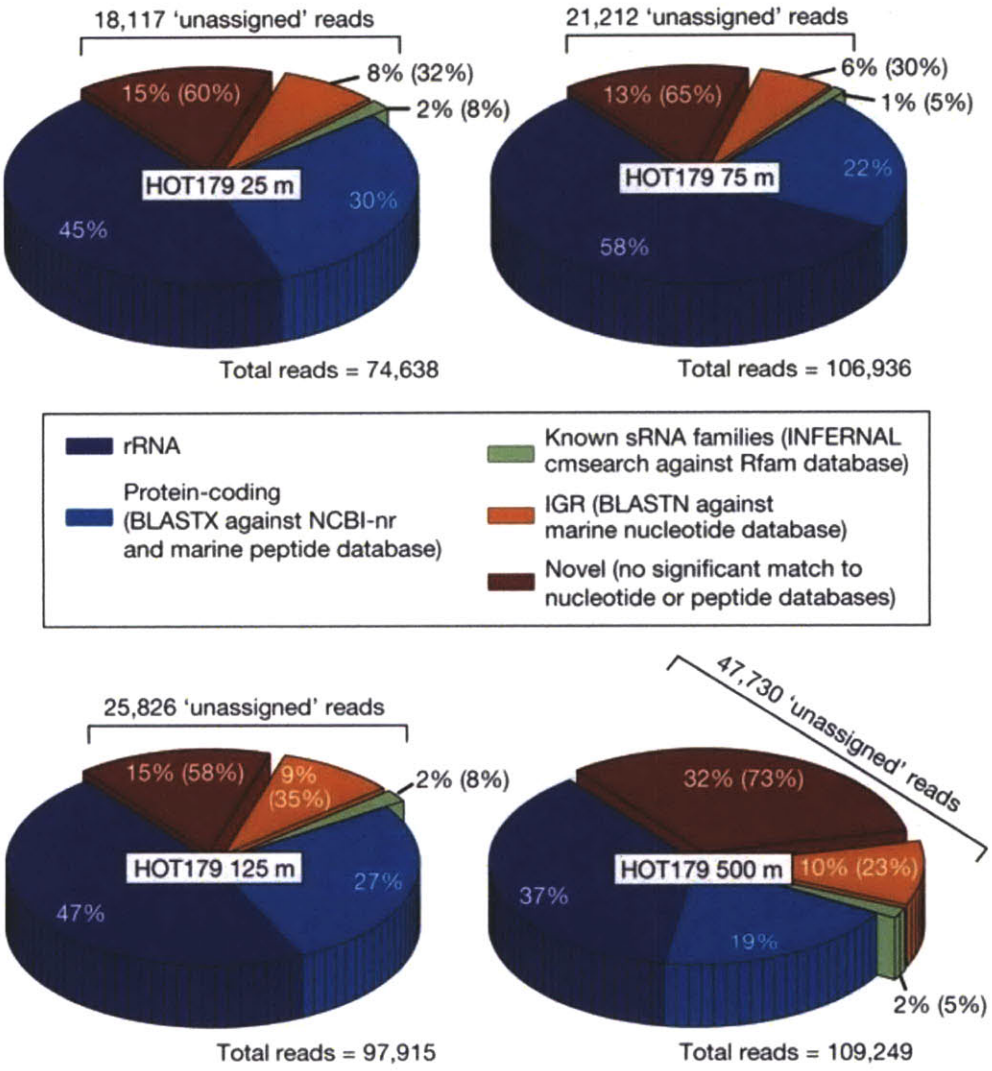
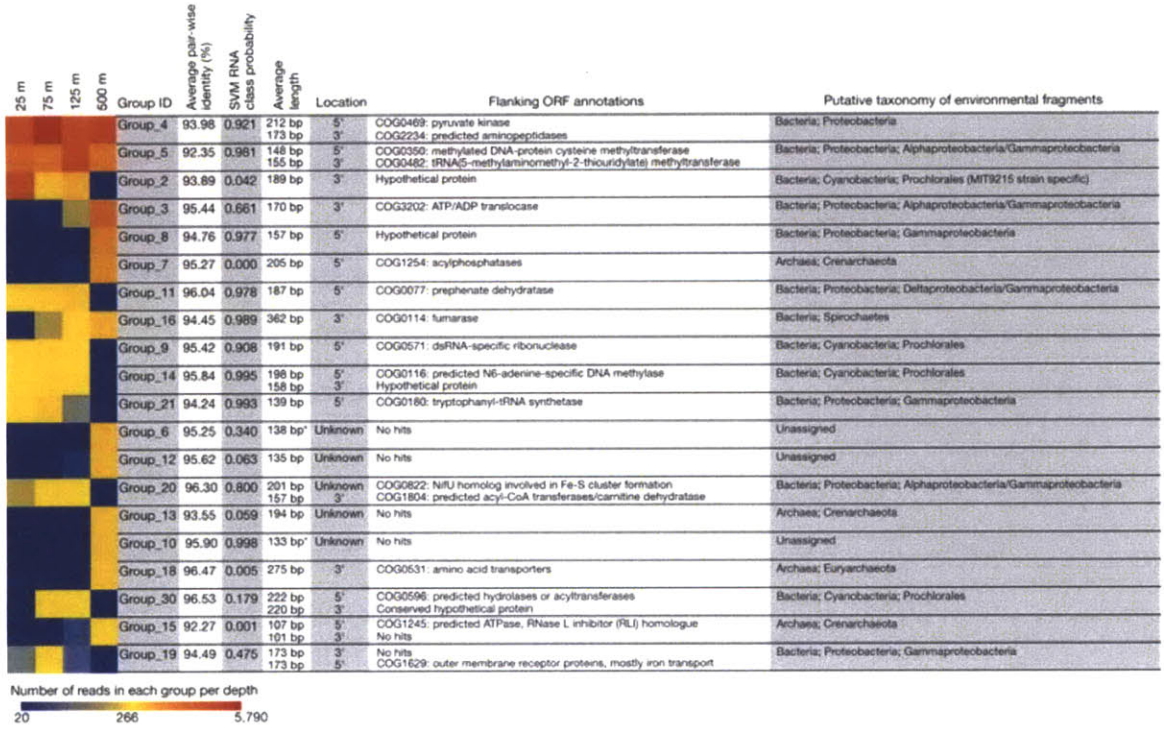


Figure 1. Inventory of RNAs in microbial community transcriptomic depth profile. The three offset slices represent reads that are not assigned to rRNA and known protein-coding genes, and are referred to as "unassigned". Numbers in parentheses represent the percentage of the total unassigned cDNA reads in each category.



Group ID	Average pair-wise identity (%)	SVM RNA class probability	Average length	Location	Flanking ORF annotations	Putative taxonomy of environmental fragments
Group_4	93.98	0.921	212 bp 173 bp	5' 3'	COG0469: pyruvate kinase COG2234: predicted aminopeptidases	Bacteria; Proteobacteria
Group_5	92.35	0.981	148 bp 155 bp	5' 3'	COG0350: methylated DNA-protein cysteine methyltransferase COG0482: tRNA(5-methylaminomethyl-2-thiouridylate) methyltransferase	Bacteria; Proteobacteria; Alphaproteobacteria/Gammaproteobacteria
Group_2	93.89	0.042	189 bp	3'	Hypothetical protein	Bacteria; Cyanobacteria; Prochlorales (MIT9215 strain specific)
Group_3	95.44	0.661	170 bp	3'	COG3202: ATP/ADP translocase	Bacteria; Proteobacteria; Alphaproteobacteria/Gammaproteobacteria
Group_8	94.76	0.977	157 bp	5'	Hypothetical protein	Bacteria; Proteobacteria; Gammaproteobacteria
Group_7	95.27	0.000	205 bp	5'	COG1254: acylphosphatases	Archaea; Crenarchaeota
Group_11	96.04	0.978	187 bp	5'	COG0077: prephenate dehydratase	Bacteria; Proteobacteria; Deltaproteobacteria/Gammaproteobacteria
Group_16	94.45	0.989	362 bp	3'	COG0114: fumarase	Bacteria; Spirochaetes
Group_9	95.42	0.908	191 bp	5'	COG0571: dsRNA-specific ribonuclease	Bacteria; Cyanobacteria; Prochlorales
Group_14	95.84	0.995	198 bp 158 bp	5' 3'	COG0116: predicted N6-adenine-specific DNA methylase Hypothetical protein	Bacteria; Cyanobacteria; Prochlorales
Group_21	94.24	0.993	139 bp	5'	COG0180: tryptophanyl-tRNA synthetase	Bacteria; Proteobacteria; Gammaproteobacteria
Group_6	95.25	0.340	138 bp*	Unknown	No hits	Unassigned
Group_12	95.62	0.063	135 bp	Unknown	No hits	Unassigned
Group_20	96.30	0.800	201 bp 157 bp	Unknown 3'	COG0822: NifU homolog involved in Fe-S cluster formation COG1804: predicted acyl-CoA transferases/carnitine dehydratase	Bacteria; Proteobacteria; Alphaproteobacteria/Gammaproteobacteria
Group_13	93.55	0.059	194 bp	Unknown	No hits	Archaea; Crenarchaeota
Group_10	95.90	0.998	133 bp*	Unknown	No hits	Unassigned
Group_18	96.47	0.005	275 bp	3'	COG0531: amino acid transporters	Archaea; Euryarchaeota
Group_30	96.53	0.179	222 bp 220 bp	5' 3'	COG0596: predicted hydrolases or acyltransferases Conserved hypothetical protein	Bacteria; Cyanobacteria; Prochlorales
Group_15	92.27	0.001	107 bp 101 bp	5' 3'	COG1245: predicted ATPase, RNase L inhibitor (RLI) homologue No hits	Archaea; Crenarchaeota
Group_19	94.49	0.475	173 bp 173 bp	3' 5'	No hits COG1629: outer membrane receptor proteins, mostly iron transport	Bacteria; Proteobacteria; Gammaproteobacteria

Figure 2. Abundance and distribution of the top twenty most abundant sRNA and psRNA groups identified in the community transcriptomic data. The twenty groups were ranked based on total abundance, and each group's depth distribution is shown in the left panel, with the number of reads in each dataset indicated by color, from high (red) to low (blue). Each group's proximity (5' or 3') to the nearest gene, annotation and putative taxonomy for that gene (where possible) are shown. The RNA-class probability values were generated with a support vector machine (SVM) learning algorithm using RNAz (Washietl et al., 2005). A complete list of sRNA and psRNA groups containing > 100 cDNA reads is provided in Supplementary Table 2.

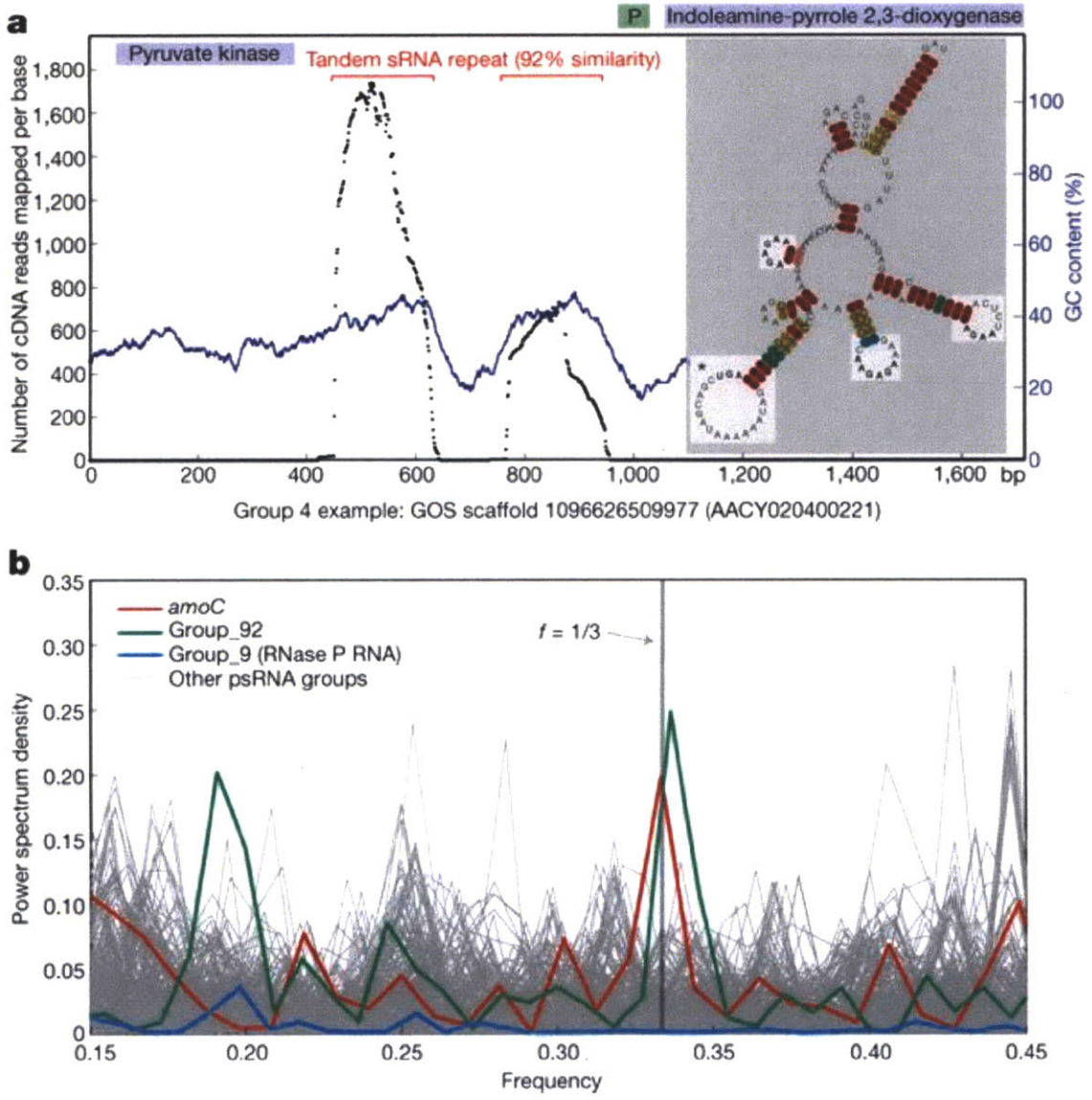


Figure 3. Characteristics of psRNA groups consistent with known sRNA. **(a)** Genomic context and features of the most abundant psRNA group, Group 4, mapped onto a *Gammaproteobacteria*-like contig from the Global Ocean Sampling (GOS) database. Sequence coverage (black dots, left axis) and reference GC content (blue dots, right axis) shown. Gene annotations are indicated along the top of the panel (upper and lower lines represent forward and reverse strands; P and T represent promoter and terminator, respectively). In the predicted structure (inset), loops containing conserved sequence motifs (in bold letters) are highlighted, and the loop marked with an asterisk contains sequences predicted to interact with 5' translation start site of a flanking gene. **(b)** Three-base periodicity analysis of multiple sequence alignments for the 66 self-clustered groups. A significant peak of power spectrum density at the frequency of 1/3 indicates 3-base periodicity in the nucleotide substitution patterns, suggesting protein-coding potentials (Ré & Pavesi, 2007). See methods for detail.

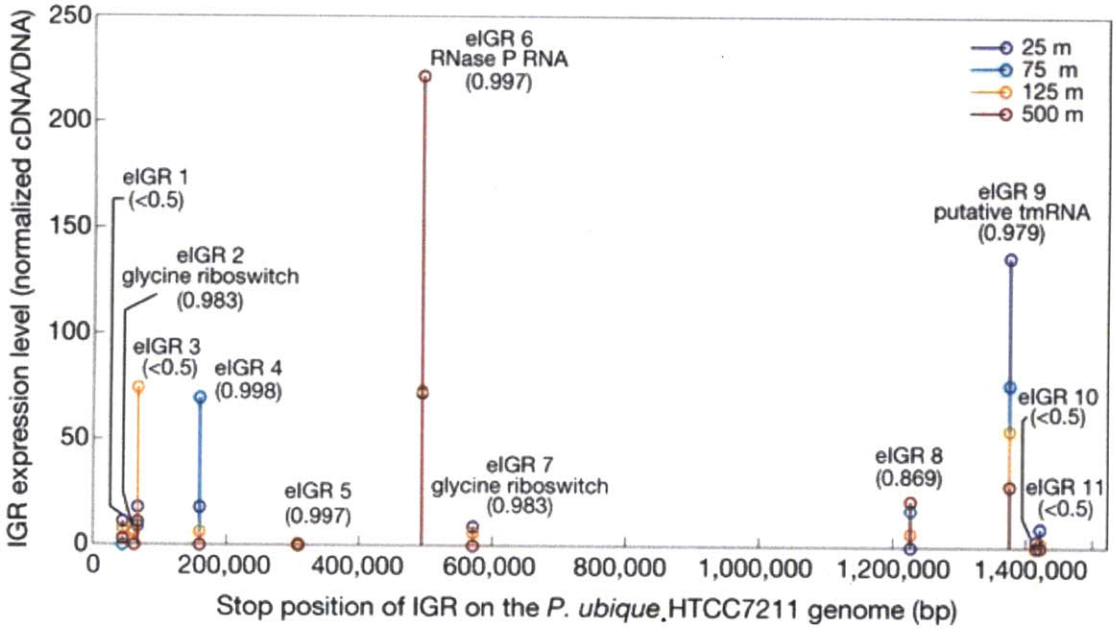


Figure 4. Normalized cDNA/DNA ratios of expressed intergenic regions (eIGR) on the *P. ubique* HTCC7211 genome, in all four depths. Since a manually curated HTCC7211 genome annotation is not yet publicly available, the genomic regions that recruited psRNAs were manually inspected and confirmed as IGRs. The values in the parentheses are RNA-class probability values generated with a support vector machine (SVM) learning algorithm using RNAz (Washietl et al., 2005).

Acknowledgements and author contributions

E.F.D. and Y.S. conceived the research, provided support, and collected the samples. Y.S. prepared samples for sequencing, and made the initial observation of sRNA sequences. E.F.D., Y.S., and G.W.T. developed the concept of the paper together. Y.S. and G.W.T. performed the data analysis. Y.S. wrote the first draft of the paper, which was completed by G.W.T. and E.F.D. together. We are grateful to the University of Hawaii HOT team, and the captain and crew of the R/V Kilo Moana for their expert assistance at sea. Thanks also to Stephan Schuster for collaboration and advice on pyrosequencing. We thank John Eppley for help with computational analyses and useful discussion, and Julia Maresca, Asuncion Martinez, Jay McCarren, and Virginia Rich for their valuable comments on this manuscript. We thank Stephen Giovannoni, Jim Tripp and Michael Schwalbach for sharing their in press manuscript on Pelagibacter riboswitches, and Stephen Giovannoni, Ulrich Stingl, the J. Craig Venter Institute, and the Gordon and Betty Moore Foundation for the genome sequence of Pelagibacter strain HTCC7211. This work was supported by the Gordon and Betty Moore Foundation, National Science Foundation Microbial Observatory Award MCB-0348001, the Department of Energy Genomics

GTL Program, and the Department of Energy Microbial Genomics Program, and an NSF Science and Technology award, C-MORE. This article is a contribution from the NSF Science and Technology Center for Microbial Oceanography: Research and Education (C-MORE).

Supplementary Information for Chapter 5

Supplementary Tables S1-S3
Supplementary Figures S1-S6

Supplementary Tables and Figures

Table S1. Distribution of known sRNA families in cDNA and DNA pyrosequence datasets (cDNA | DNA). The shaded rows represent sRNAs that were found in both datasets, and are ranked by the ratio of total counts in cDNA dataset to the total counts in DNA dataset.

Rfam id;annotation	Function	Total # of reads	# of reads per depth			
			25m	75m	125m	500m
RF00162;SAM	Riboswitch; methionine/cysteine biosynthesis	4 | 0	0 | 0	0 | 0	0 | 0	4 | 0
RF00029;Intron_gpII	Self-splicing ribozyme	2 | 0	0 | 0	0 | 0	0 | 0	2 | 0
RF00016;SNORD14	Cleavage of eukaryotic precursor rRNA	2 | 0	0 | 0	1 | 0	1 | 0	0 | 0
RF00169;SRP_bact	Translation and targeting of proteins to cell membranes	474 | 30	101 | 9	94 | 11	148 | 4	131 | 6
RF00004;U2	Pre-mRNA splicing in eukaryotes	14 | 1	1 | 1	5 | 0	1 | 0	7 | 0
RF00010;RNaseP_bact_a	Generation of mature tRNA	833 | 63	238 | 27	267 | 21	194 | 12	134 | 3
RF00023;tmRNA	Rescue of stalled ribosomes; cell cycle regulation	1961 | 200	242 | 38	413 | 49	539 | 50	766 | 63
RF00013;6S	Gene regulation during stationary phase	71 | 18	12 | 7	23 | 10	33 | 1	3 | 0
RF00504;Glycine	Riboswitch; glycine metabolism	29 | 17	17 | 3	6 | 3	5 | 7	1 | 4
RF00005;tRNA	Protein synthesis	1036 | 874	175 | 214	138 | 259	490 | 205	232 | 196
RF00059;TPP	Riboswitch; gene regulation	7 | 15	1 | 2	0 | 4	4 | 0	2 | 9
RF00174;Cobalamin	Riboswitch; gene regulation	2 | 6	1 | 2	1 | 1	0 | 1	0 | 2
RF00017;SRP_euk_arch	Translation and targeting of proteins to cell membranes	4 | 43	1 | 19	1 | 5	1 | 3	1 | 16
RF00519;suhB	Putative sRNA with unknown function	0 | 2	0 | 0	0 | 0	0 | 0	0 | 2
RF00066;U7	Pre-mRNA splicing in eukaryotes	0 | 5	0 | 1	0 | 0	0 | 3	0 | 1
RF00582;SCARNA14	Small nuclear RNA in eukaryotes	0 | 8	0 | 3	0 | 0	0 | 2	0 | 3
RF00521;SAM_alpha	Riboswitch; methionine biosynthesis in Alphaproteobacteria	0 | 10	0 | 2	0 | 0	0 | 2	0 | 6

Table S2 (next page). Mapping of sRNA and psRNA groups (represented by more than 100 cDNA reads) onto environmental nucleotide sequences. Reads from each group were compared to the NCBI env-nt datasbase, as well as a database of marine-specific metagenomic sequences using BLASTN. Reads were assigned to the top blast hit above cutoff scores (if multiple top hits were obtained, all were counted). sRNA and psRNA groups that either did not have significant matches in available databases or the matched sequences did not contain predicted protein coding genes were not included (17 groups out of 66). sRNA groups that can be confidently assigned to Rfam sRNA families are marked with an asterisk. Using the frequency of reads mapping to each of these environmental fragments, it was possible to determine the attributes of these groups binned by nearest flanking protein-coding gene (COG annotation) in each group: predicted location [5', 3', NA (not assigned if sRNA is not flanked by one ORF on each side)], the number of different environmental fragments hit by reads in each COG bin in each group, and the distribution of hits with depth. The average length and coverage (the number of times any base in the region is sampled) of psRNAs was also calculated, and putative taxonomic origin of psRNAs was predicted based on the taxonomy of the flanking coding regions.

Group ID	Location	SVM RNA probability	Avg. length (bp)	Avg. coverage (per bp)	COG annotation of flanking gene	Putative Taxonomy	# of env-nt fragments hit	Total number of reads hitting environmental fagments per denth			
								25m	75m	125m	500m
Group_4	5' NA		212	93.30	COG0469: Pyruvate kinase	Bacteria; Proteobacteria; Gammaproteobacteria Bacteria; Spirochaetes; Spirochaetales	30	3564	5315	3061	2
Group_4	NA 3'		173	137.21	COG2234: Predicted aminopeptidases	Bacteria; Proteobacteria; Alphaproteobacteria	14	2088	2453	1657	0
Group_4	3' NA		154	74.38	COG3816: Uncharacterized protein conserved in bacteria	Bacteria; Proteobacteria; Gammaproteobacteria	20	740	1290	1249	1350
Group_4	3' NA		172	73.29	COG2225: Malate synthase	Bacteria; Proteobacteria; Alphaproteobacteria Bacteria; environmental samples;	18	1434	1484	1408	0
Group_4	NA 3'		139	60.11	COG0004: Ammonia permease	Bacteria; Proteobacteria; Gammaproteobacteria Bacteria; environmental samples;	6	544	524	435	0
Group_4	5'		161	87.96	COG1629: Outer membrane receptor proteins, mostly Fe transport	Bacteria; Proteobacteria; Gammaproteobacteria	4	528	541	396	0
Group_4	NA		161	62.45	COG0644: Dehydrogenases (flavoproteins)	Bacteria; Proteobacteria; Gammaproteobacteria	5	87	132	195	631
Group_4	5' NA		154	62.60	COG0654: 2-polyprenyl-6-methoxyphenol hydroxylase and related FAD-dependent oxidoreductases	Bacteria; Actinobacteria; Actinobacteridae	4	296	448	238	0
Group_4	NA	0.95	185	126.56	COG1028: 8 Dehydrogenases with different specificities (related to short-chain alcohol dehydrogenases)	Bacteria; Proteobacteria; Deltaproteobacteria	2	466	452	42	0
Group_4	5' NA		165	42.13	COG4146: Predicted symporter	Bacteria; Proteobacteria; Gammaproteobacteria	6	322	368	234	4
Group_4	5' NA 3'		160	34.69	COG4667: Predicted esterase of the alpha-beta hydrolase superfamily	Bacteria; Proteobacteria; Gammaproteobacteria	8	162	279	216	228
Group_4	3'		133	49.40	COG0800: 2-keto-3-deoxy-6-phosphogluconate aldolase	Bacteria; Proteobacteria; Epsilonproteobacteria	6	224	204	165	280
Group_4	5' NA		167	25.06	COG0516: IMP dehydrogenase/GMP reductase	Bacteria; Proteobacteria; Gammaproteobacteria Eukaryota; Choanoflagellida; Codonosigidae	3	104	164	343	0
Group_4	NA 3'		153	37.88	COG3250: Beta-galactosidase/beta-glucuronidase	Bacteria; Proteobacteria; Gammaproteobacteria	4	188	189	179	19
Group_4	NA		181	35.27	COG0072: Phenylalanyl-tRNA synthetase beta subunit	Bacteria; Proteobacteria; Gammaproteobacteria	1	60	78	194	0
Group_4	3'		149	76.98	COG1609: Transcriptional regulators	Bacteria; Proteobacteria; Gammaproteobacteria	1	112	101	81	1
Group_4	5'		149	34.01	COG0837: Glucokinase	Bacteria; Proteobacteria; Alphaproteobacteria	2	2	2	8	239
Group_4	5' NA		172	18.24	COG2609: Pyruvate dehydrogenase complex, dehydrogenase (E1) component	Bacteria; Proteobacteria; Gammaproteobacteria	2	43	37	78	0
Group_4	5'		173	12.44	COG0119: Isopropylmalate/homocitrate/citramalate synthases	Bacteria; Proteobacteria; Gammaproteobacteria	2	45	42	39	2
Group_4	NA		98	23.19	COG0492: Thioredoxin reductase	Bacteria; environmental samples;	1	22	34	22	0
Group_4	NA		179	14.24	COG0508: Pyruvate/2-oxoglutarate dehydrogenase complex, dihydrolipoamide acyltransferase (E2) component, and related enzymes	Bacteria; Proteobacteria; Gammaproteobacteria	1	19	18	30	0
Group_5	5' NA		148	39.77	COG0350: Methylated DNA-protein cysteine methyltransferase	Bacteria; Proteobacteria; Alphaproteobacteria (Pelagibacter)	157	6958	4320	7252	1106
Group_5	NA 3'		155	29.75	COG0482: Predicted tRNA(5-methylaminomethyl-2-thiouridylate) methyltransferase, contains the PP-loop ATPase domain	Bacteria; Proteobacteria; Alphaproteobacteria (Pelagibacter)	28	734	613	1342	636
Group_5	3' NA		145	33.41	COG4241: Predicted membrane protein	Bacteria; Proteobacteria; Gammaproteobacteria	8	147	110	219	4
Group_5	3'		176	42.15	COG4781: Membrane domain of membrane-anchored glycerophosphoryl diester phosphodiesterase	Bacteria; Proteobacteria; Gammaproteobacteria	2	68	64	99	2
Group_5	NA		171	35.99	COG0582: Integrase	Bacteria; Proteobacteria; Alphaproteobacteria (Pelagibacter)	1	87	52	86	0
Group_5	3'	0.98	137	53.95	COG0833: Amino acid transporters	Bacteria; Proteobacteria; Gammaproteobacteria	1	36	33	59	1
Group_5	NA		166	14.19	COG0477: 77 Permeases of the major facilitator superfamily	Bacteria; Proteobacteria; Alphaproteobacteria (Pelagibacter)	1	0	0	1	126
Group_5	3'		65	12.86	COG0697: 7 Permeases of the drug/metabolite transporter (DMT) superfamily	Bacteria; Proteobacteria; Alphaproteobacteria (Pelagibacter)	1	12	16	35	0
Group_5	3'		149	42.80	COG0451: Nucleoside-diphosphate-sugar epimerases	Bacteria; Bacteroidetes; Sphingobacteria	1	16	18	26	0
Group_5	3'		153	25.93	COG1530: Ribonucleases G and E	Bacteria; Proteobacteria; Alphaproteobacteria (Pelagibacter)	1	10	5	36	2
Group_5	5'		158	13.24	COG2133: Glucose/sorbosone dehydrogenases	Bacteria; Proteobacteria; Betaproteobacteria	1	8	9	21	0
Group_5	3'		121	12.22	COG2721: Altronate dehydratase	Bacteria; Proteobacteria; Alphaproteobacteria (Pelagibacter)	1	4	8	13	4
Group_2	NA 3' 5'	0.02	189	63.21	no hit: unknown	Bacteria; Cyanobacteria; Prochlorales	71	12107	2601	4615	18

Group_3	NA 3'		170	45.91	COG3202: ATP/ADP translocase	Bacteria; Proteobacteria; Alphaproteobacteria	9	11	128	681	1922
Group_3	NA	0.68	168	59.04	COG1540: Uncharacterized proteins, homologs of lactam utilization protein B	Bacteria; Proteobacteria; Gammaproteobacteria Bacteria; Acidobacteria; Solibacteres Bacteria; Deinococcus-Thermus; Deinococci	5	6	72	361	1641
Group_3	3'		143	40.28	no hit: unknown	Bacteria; Proteobacteria; Gammaproteobacteria Bacteria; Proteobacteria; Alphaproteobacteria Bacteria; Proteobacteria; Alphaproteobacteria	5	6	65	322	352
Group_8	5'	0.97	157	133.84	no hit: unknown	Bacteria; Proteobacteria; Gammaproteobacteria	1	0	0	0	511
Group_7	NA		205	54.59	COG1254: Acylphosphatases	Archaea; Crenarchaeota; Thermoprotei	6	0	0	0	1213
Group_7	NA 5'	0.0005	234	38.77	no hit: unknown	Archaea; Crenarchaeota; Thermoprotei	8	0	3	226	706
Group_7	NA		270	61.72	COG0037: Predicted ATPase of the PP-loop superfamily implicated in cell cycle control	Archaea; Crenarchaeota; Thermoprotei	2	0	1	47	382
Group_11	5' NA	0.98	187	62.08	COG0077: Prephenate dehydratase	Bacteria; Proteobacteria; Deltaproteobacteria	16	1136	1478	1798	12
Group_11	NA		151	37.59	COG0300: Short-chain dehydrogenases of various substrate specificities	Bacteria; Proteobacteria; Gammaproteobacteria Bacteria; Actinobacteria; Actinobacteridae	2	124	158	46	2
Group_16	3' NA	0.99	362	25.07	COG0114: Fumarase	Bacteria; Spirochaetes; Spirochaetales Bacteria; Proteobacteria; Deltaproteobacteria Eukaryota; Metazoa; Chordata Eukaryota; Fungi; Dikarya	12	67	519	1216	304
Group_16	NA		166	47.36	COG1530: Ribonucleases G and E	Eukaryota; Metazoa; Chordata	1	0	4	103	8
Group_9*	NA 5'	0.90	191	17.71	COG0571: dsRNA-specific ribonuclease	Bacteria; Cyanobacteria; Prochlorales	29	793	729	637	4
Group_9*	NA 5'		172	20.70	no hit: unknown	Bacteria; Cyanobacteria; Prochlorales	16	550	507	425	4
Group_14	NA 5'		198	33.95	COG0116: Predicted N6-adenine-specific DNA methylase	Bacteria; Cyanobacteria; Prochlorales	17	1012	1136	833	7
Group_14	NA 3' 5'	1.00	158	37.34	no hit: unknown	Bacteria; Cyanobacteria; Prochlorales	16	811	828	684	8
Group_14	NA 3'		165	20.94	COG0667: Predicted oxidoreductases (related to aryl-alcohol dehydrogenases)	Bacteria; Cyanobacteria; Prochlorales	4	128	162	116	2
Group_21	5' NA	0.99	139	45.30	COG0180: Tryptophanyl-tRNA synthetase	Bacteria; Proteobacteria; Gammaproteobacteria	12	721	885	301	0
Group_21	5'		141	31.58	no hit: unknown	Bacteria; Proteobacteria; Gammaproteobacteria	1	60	63	12	0
Group_12	NA	0.06	135	31.79	no hit: unknown	unknown	4	2	2	105	762
Group_20	NA		201	33.77	COG0822: NifU homolog involved in Fe-S cluster formation	Bacteria; Proteobacteria; Alphaproteobacteria	6	176	240	392	0
Group_20	5' NA	0.78	158	57.05	no hit: unknown	Bacteria; Proteobacteria; Gammaproteobacteria	5	173	182	246	4
Group_20	3' NA		215	22.65	COG0441: Threonyl-tRNA synthetase	Bacteria; Proteobacteria; Gammaproteobacteria	3	63	83	156	0
Group_20	3' NA		157	24.62	COG1804: Predicted acyl-CoA transferases/carnitine dehydratase	Bacteria; Proteobacteria; Alphaproteobacteria	4	32	70	138	0
Group_13	NA	0.06	194	83.91	no hit: unknown	Archaea; Crenarchaeota; Thermoprotei	3	0	0	0	783
Group_18	NA	0.004	93	59.10	no hit: unknown	unknown	4	0	0	0	530
Group_18	3'		275	114.19	COG0531: Amino acid transporters	Archaea; Euryarchaeota; Archaeoglobi	1	0	0	0	423
Group_30	NA 5'	0.20	222	27.96	COG0596: Predicted hydrolases or acyltransferases (alpha/beta hydrolase superfamily)	Bacteria; Cyanobacteria; Prochlorales	18	38	1132	1007	4
Group_30	NA 3'		220	38.89	no hit: unknown	Bacteria; Cyanobacteria; Prochlorales	13	36	1129	926	6
Group_15	3' 5' NA	0.001	101	26.88	no hit: unknown	Archaea; Crenarchaeota; Thermoprotei Bacteria; Tenericutes; Mollicutes	11	0	0	332	463
Group_15	5' NA		107	24.22	COG1245: Predicted ATPase, RNase L inhibitor (RLI) homolog	Archaea; Crenarchaeota; Thermoprotei	6	0	0	332	329
Group_19	3' NA	0.46	173	39.48	no hit: unknown	unknown	4	191	317	168	0
Group_19	5' NA		173	32.99	COG1629: Outer membrane receptor proteins, mostly Fe transport	Bacteria; Proteobacteria; Gammaproteobacteria	2	103	143	97	0
Group_29	3'	0.89	170	26.79	no hit: unknown	Bacteria; Bacteroidetes; Flavobacteria Bacteria; Proteobacteria; Gammaproteobacteria	8	40	360	464	74
Group_29	NA		170	10.34	COG1196: Chromosome segregation ATPases	Bacteria; Bacteroidetes; Flavobacteria	2	2	52	60	2
Group_48	NA 3' 5'	0.01	171	25.81	no hit: unknown	Bacteria; Cyanobacteria; Prochlorales	18	58	533	1302	0
Group_32	NA 3' 5'	0.95	139	50.17	no hit: unknown	unknown	8	12	202	536	426
Group_22	3'	0.83	121	80.25	no hit: unknown	unknown	4	23	225	203	0

Group_25	5' NA		83	34.20	COG2124: Cytochrome P450	Bacteria; Proteobacteria; Gammaproteobacteria	8	16	550	114	0
Group_25	5'	0.92	75	17.48	COG0258: 5'-3' exonuclease (including N-terminal domain of PolI)	Bacteria; Actinobacteria; Actinobacteridae Bacteria; Proteobacteria; Deltaproteobacteria Bacteria; Acidobacteria; Acidobacteriales	4	8	143	52	0
Group_25	NA		75	40.63	COG1247: Sortase and related acyltransferases	Bacteria; Firmicutes; Bacillales	1	2	64	14	0
Group_23	ORF	0.11	129	60.83	no hit: unknown	unknown	6	19	493	251	0
Group_56*	5'	0.92	236	30.00	no hit: unknown	Bacteria; Firmicutes; Clostridia	2	12	14	18	150
Group_56*	NA		152	18.58	COG1351: Predicted alternative thymidylate synthase	Bacteria; Proteobacteria; Alphaproteobacteria (Pelagibacter)	1	0	5	5	56
Group_50	NA	0.99	320	27.56	no hit: unknown	unknown	3	2	27	246	50
Group_50	5' NA		143	16.09	COG1024: Enoyl-CoA hydratase/carnithine racemase	Bacteria; Actinobacteria; Actinobacteridae	4	0	12	238	8
Group_24	NA 3'	0.98	240	71.96	no hit: unknown	Bacteria; Actinobacteria; Actinobacteridae	2	0	0	0	434
Group_17	3' NA	0.98	108	22.16	no hit: unknown	Bacteria; Cyanobacteria; Prochlorales	17	14	37	1632	2
Group_45	NA	0.99	338	19.66	COG0369: Sulfite reductase, alpha subunit (flavoprotein)	Archaea; Euryarchaeota; Marine Group II	2	42	106	246	2
Group_27	NA		128	55.59	COG0072: Phenylalanyl-tRNA synthetase beta subunit	Archaea; Crenarchaeota; Thermoprotei	2	0	0	0	272
Group_27	3'	0.18	108	29.57	no hit: unknown	unknown	4	0	0	0	199
Group_27	NA		137	11.50	COG2947: Uncharacterized conserved protein	Archaea; Crenarchaeota; Thermoprotei	1	0	0	0	25
Group_26*	3' NA	0.95	88	42.77	COG0206: Cell division GTPase	Bacteria; Cyanobacteria; Prochlorales	9	321	225	284	0
Group_26*	3'		90	67.03	no hit: unknown	Bacteria; Cyanobacteria; Prochlorales	2	102	68	86	0
Group_52	5'	0.27	211	13.62	COG1475: Predicted transcriptional regulators	Bacteria; Firmicutes; Clostridia	2	12	12	38	8
Group_52	ORF		53	19.48	no hit: unknown	unknown	2	4	12	18	0
Group_34	5' NA	0.94	272	26.91	COG2062: Phosphohistidine phosphatase SixA	Bacteria; Proteobacteria; Alphaproteobacteria	4	81	402	156	0
Group_34	NA		234	17.19	COG0673: Predicted dehydrogenases and related proteins	Bacteria; Proteobacteria; Alphaproteobacteria	2	18	206	66	0
Group_58	5' NA	0.13	302	19.73	no hit: unknown	Bacteria; Proteobacteria; Deltaproteobacteria	2	50	87	96	5
Group_58	NA		301	10.32	COG0290: Translation initiation factor 3 (IF-3)	Bacteria; environmental samples;	1	26	45	47	3
Group_43	NA	0.05	270	52.91	no hit: unknown	unknown	1	0	0	0	261
Group_44	NA	0.07	69	10.69	COG1976: Translation initiation factor 6 (eIF-6)	Archaea; Euryarchaeota; Marine Group II	2	14	4	50	41
Group_59	5' NA	1.00	141	42.15	COG0180: Tryptophanyl-tRNA synthetase	Bacteria; Proteobacteria; Gammaproteobacteria	11	321	649	508	0
Group_67	5' 3'	0.95	92	14.28	no hit: unknown	Bacteria; Proteobacteria; Alphaproteobacteria	4	3	55	80	4
Group_67	5'		97	14.23	COG0590: Cytosine/adenosine deaminases	Bacteria; Proteobacteria; Alphaproteobacteria	2	2	32	56	0
Group_66	NA	0.04	145	26.28	no hit: unknown	Bacteria; Proteobacteria; Betaproteobacteria	1	0	0	0	45
Group_28	NA	1.00	298	12.36	COG1028: 8 Dehydrogenases with different specificities (related to short-chain alcohol dehydrogenases)	Bacteria; Proteobacteria; Betaproteobacteria	1	7	78	30	0
Group_47*	3' NA	0.75	63	21.55	COG0054: Riboflavin synthase beta-chain	Bacteria; Cyanobacteria; Prochlorales	4	48	4	194	2
Group_47*	5' NA		63	19.08	no hit: unknown	Bacteria; Cyanobacteria; Prochlorales	2	24	2	97	1
Group_46	5'	0.98	370	14.62	COG2838: Monomeric isocitrate dehydrogenase	Bacteria; Bacteroidetes; Flavobacteria	2	0	2	58	122
Group_49*	NA 3'	0.90	62	11.59	COG1523: Type II secretory pathway, pullulanase PulA and related glycosidases	Bacteria; Cyanobacteria; Prochlorales	3	10	27	74	0
Group_49*	5'		62	11.42	no hit: unknown	unknown	1	5	9	26	0
Group_64	NA	0.10	123	12.51	no hit: unknown	Bacteria; Proteobacteria; Gammaproteobacteria	2	4	74	28	0
Group_36	ORF	0.02	85	39.75	no hit: unknown	unknown	1	31	4	8	10
Group_54	3'	0.30	271	19.39	COG0405: Gamma-glutamyltransferase	Bacteria; Proteobacteria; Alphaproteobacteria	2	16	152	106	0
Group_90	NA	0.99	229	11.74	no hit: unknown	Bacteria; Bacteroidetes; Flavobacteria	4	8	146	98	0
Group_90	NA		112	12.05	COG1741: Pirin-related protein	Eukaryota; Alveolata; Ciliophora	2	2	16	34	0
Group_39*	5'	0.01	296	21.82	COG2001: Uncharacterized protein conserved in bacteria	Bacteria; Proteobacteria; Gammaproteobacteria	2	2	0	0	222
Group_94	5' NA	0.07	231	20.60	no hit: unknown	Bacteria; Proteobacteria; Gammaproteobacteria	8	211	180	279	8
Group_94	5' NA		333	20.79	COG0579: Predicted dehydrogenase	Eukaryota; Metazoa; Chordata	4	101	110	168	2

Table S3. Features of expressed IGRs of *Candidatus Pelagibacter ubique* HTCC7211 genome. eIGR represents expressed ingergenic region.

sRNA detected	genomic location	genomic context [a]	function	adjacent ORFs	SVM RNA probability [b]
eIGR #1	IGR [44547..44776]	← →	unknown	dTDP glucose 4, 6-dehydratase; 23S rRNA gene	< 0.5
eIGR #2	IGR [61849..62122]	← →	glycine riboswitch	acetyl-CoA carboxylase, carboxyl transferase; malate synthase	0.983
eIGR #3	IGR [66868..67008]	← ←	unknown	DNA-directed DNA polymerase gamma/tau subunit; prephenate dehydratase	< 0.5
eIGR #4	IGR [159922..160288]	→ ←	unknown	CarD-like transcriptional regulator family; long-chain-fatty-acid--CoA ligase	0.998
sRNA #5	IGR [307432..307592]	← →	unknown	diaminopimelate epimerase; signal recognition particle protein	0.997
eIGR #6	IGR [493441..494095]	→ →	RNase P	N-acetylmuramoyl-L-alanine amidase YbjR; cell division protein MraZ	0.997
eIGR #7	IGR [570785..571078]	→ →	glycine riboswitch	trap dicarboxylate transporter; glycine cleavage system T protein	0.983
eIGR #8	IGR [1226239..1226509]	→ →	unknown	conserved hypothetical; ammonium transporter	0.869
eIGR #9	IGR [1375367..1375674]	← ←	putative tmRNA	Predicted alternative thymidylate synthase; pyruvate, phosphate dikinase	0.979
eIGR #10	IGR [1415400..1415665]	→ →	unknown	inositol monophosphatase family protein; Uncharacterized protein conserved in bacteria	< 0.5
eIGR #11	IGR [1421234..1421469]	← ←	unknown	ADP-L-glycero-D-mannoheptose-6-epimerase; Chain length determinant protein	< 0.5

[a] The arrows represent the gene orientation of the flanking ORFs

[b] The probability values were predicted by comparing structure conservation of IGRs of three Pelagibacter genomes

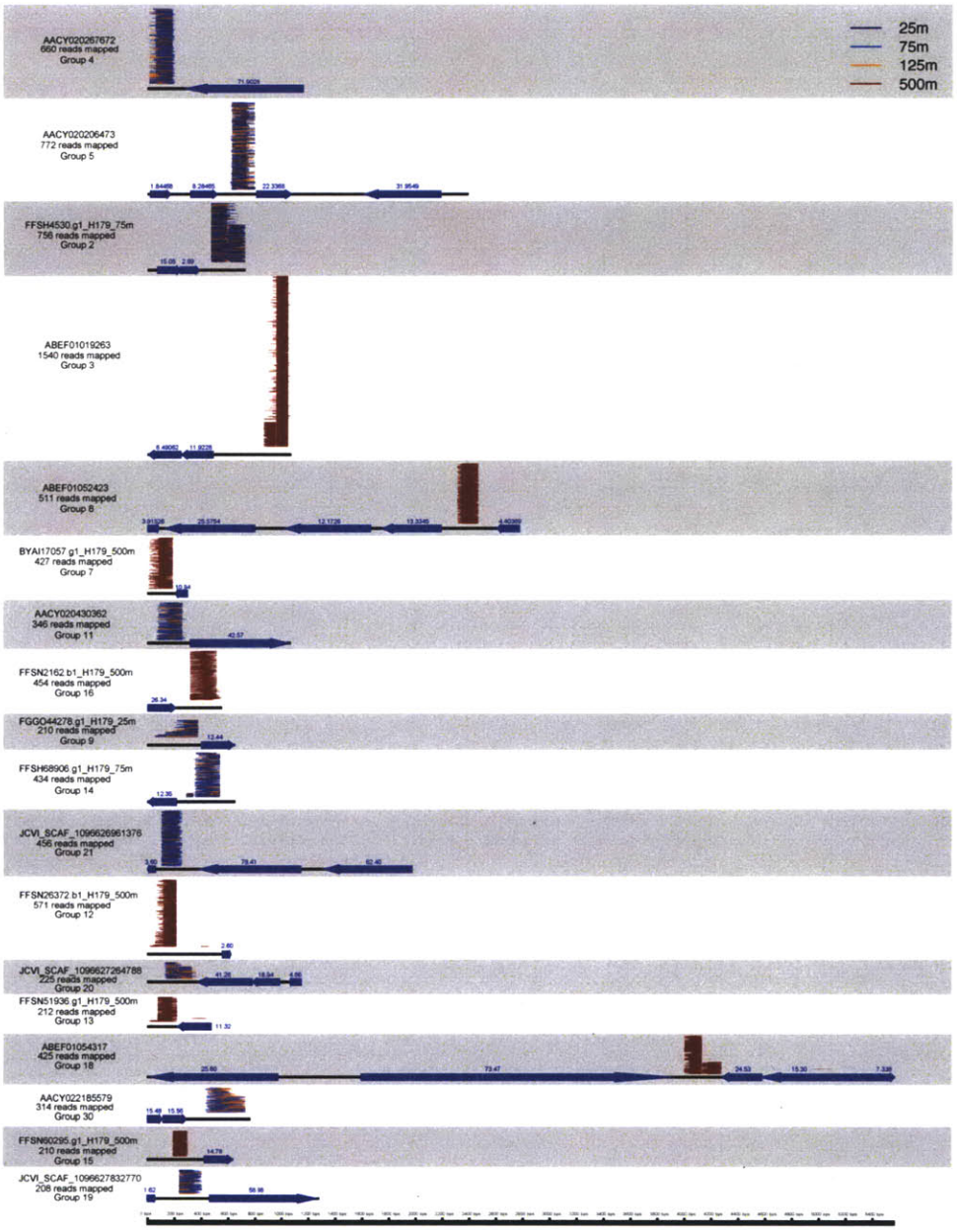


Figure S1. Mapping of cDNA reads from the most abundant groups (as shown in Fig. 2) to predicted intergenic regions on environmental genomic fragments. All reads were mapped with ≥ 85% sequence identity over 90% of the length. Two novel psRNA groups (Group 6 and 10) are not shown due to lack of reference genomic sequences. The environmental genome fragments were taken from three sources: env-nt from NCBI, GOS peptides (read ID starting with "JCVI"), and fosmid-end or shotgun sequences (read ID containing "H179"). Open reading frames (ORFs) on these environmental genomic fragments were predicted using MetaGene and estimated gene values (confidence scores) for the predictions appear above each ORF. Only ORFs with estimated gene scores > 1 are considered significant.

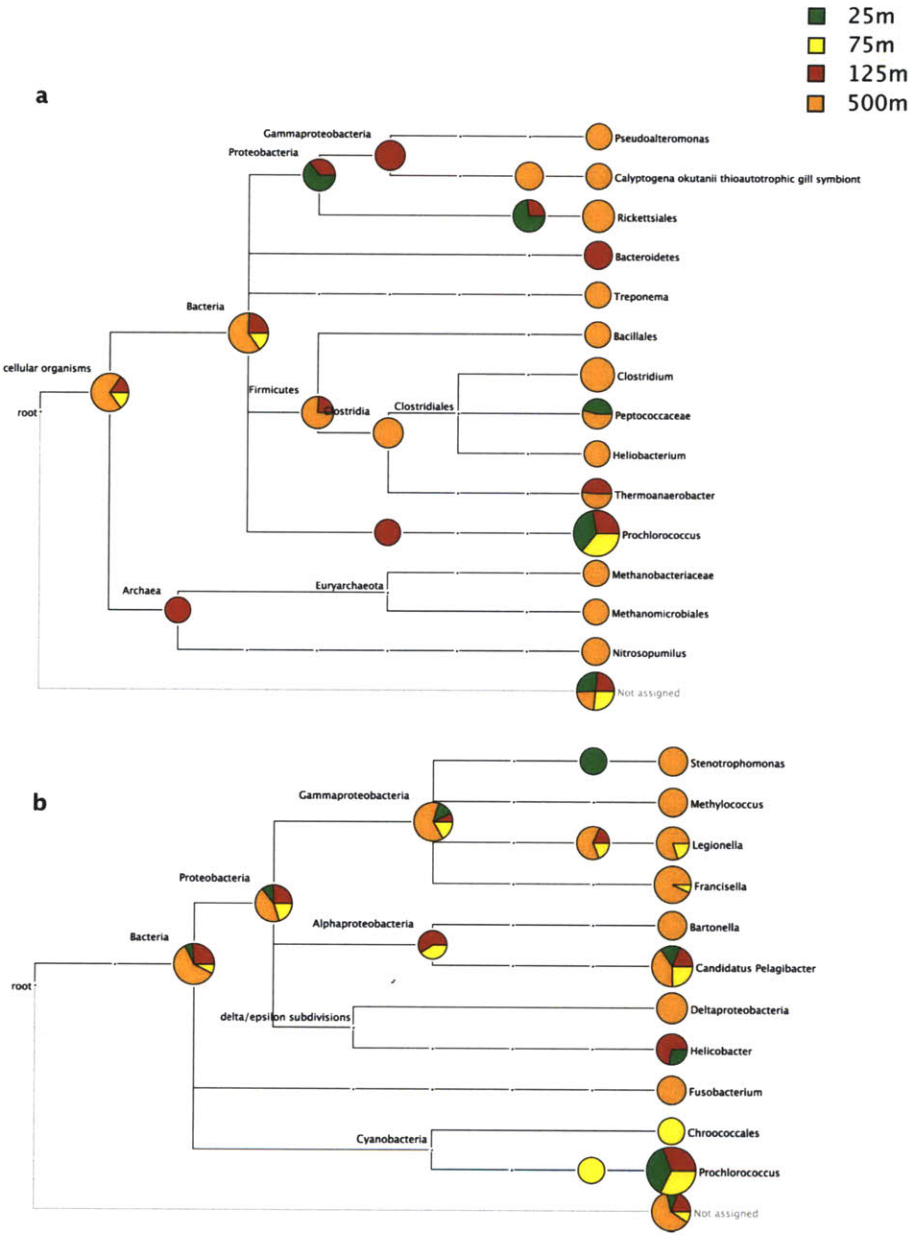


Figure S2. Putative taxonomy assignment of cDNA reads assigned to known sRNA families. The taxonomy assignment was performed using MEGAN with the default parameters, based on the output of BLASTN against NCBI-nt database. For each individual cDNA read, the taxonomic classifications of all matching sequences were analyzed to find the node of lowest common ancestor. The trees were collapsed at Genus level. **(a)** Signal Recognition Particle (SRP) RNA. **(b)** RNase P RNA. Out of four types (Type A and B for bacteria and Type A and M for archaea), only Type A bacterial RNaseP RNA was found in our transcriptomic datasets.

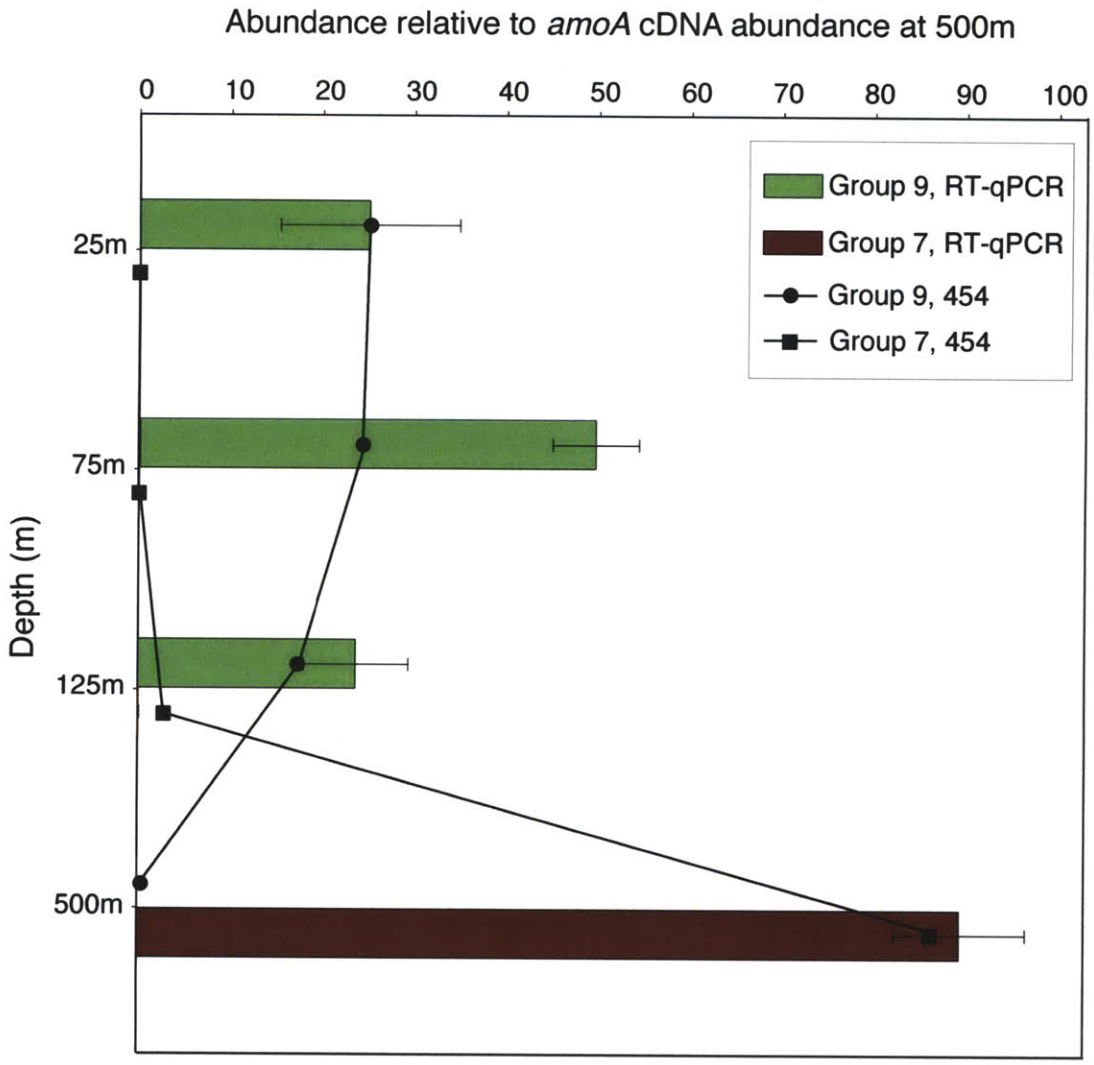


Figure S3. Verification of the abundance and depth-dependent distribution of psRNA Group 7 and sRNA Group 9 (RNase P RNA) using RT-qPCR. The bars represent the abundance of these groups relative to crenarchaeal *amoA* transcript in the 500m sample measured by RT-qPCR. The lines with markers (square: Group 7; circle: Group 9) represent the number of 454 reads assigned to each group, normalized to the corresponding gene length and the number of cDNA reads assigned to crenarchaeal *amoA* in the 500m sample.

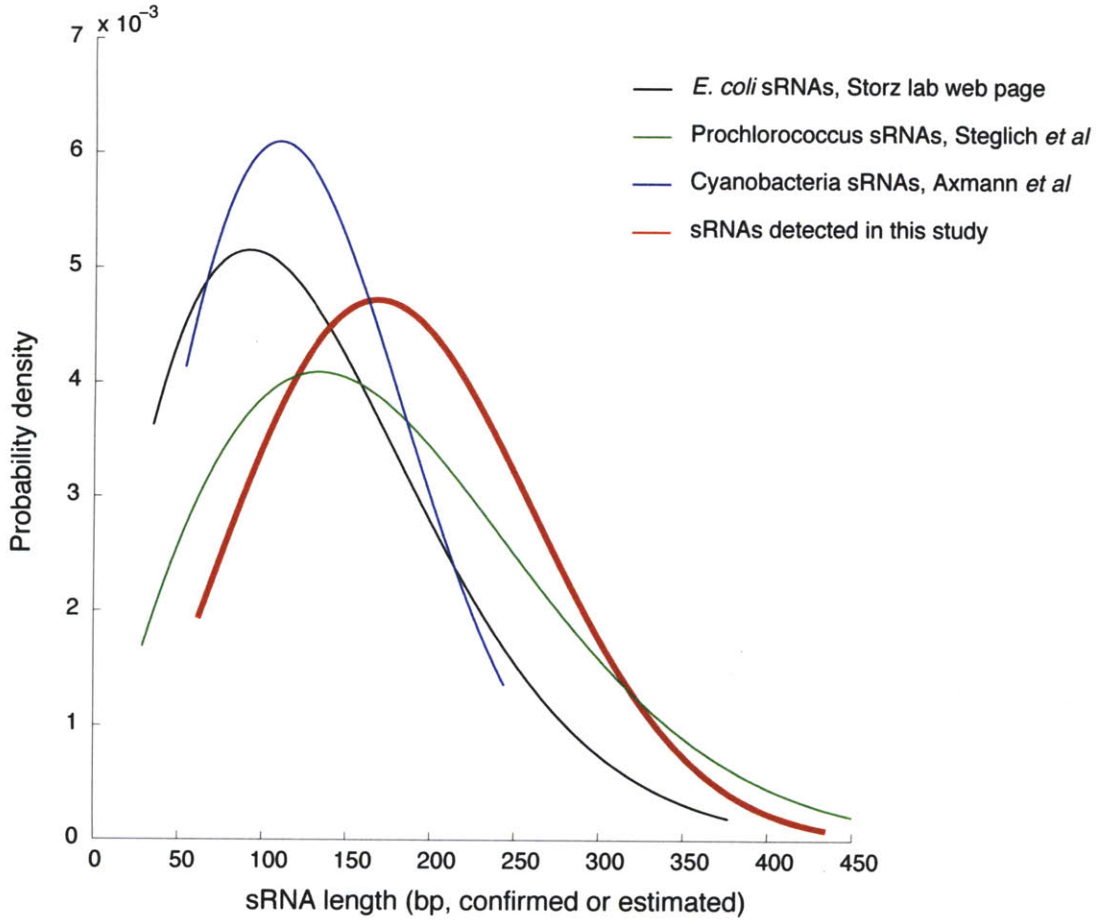


Figure S4. Comparison of the size distribution of psRNA groups identified in this study with that of known sRNAs from model organisms. The length of the psRNAs detected in this study was estimated as described in the Methods. The length of sRNAs reported in the model organisms (Axmann et al., 2005; Steglich et al., 2008; Storz, Altuvia & Wassarman, 2005) was either computationally predicted or experimentally verified.

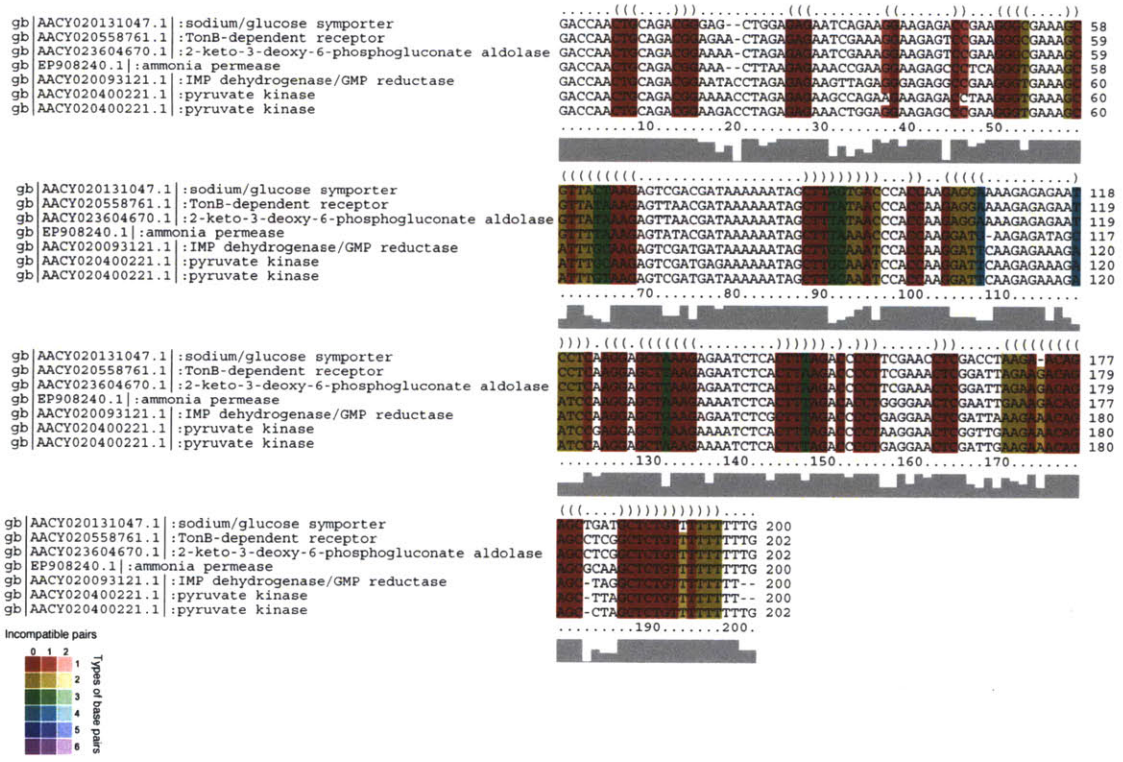


Figure S5. Multiple sequence alignment of Group 4 psRNAs. The full-length psRNA sequences were extracted from metagenomic contig sequences with different genomic context, and the nearest flanking gene was listed for each metagenomic contig. The genomic fragment AACY020400221 contains tandem copies of Group 4 psRNA, both of which are shown in the alignment. The alignment is color-coded according to the different types of base pairs and the amount of compensatory and incompatible base changes in the corresponding alignment columns (see color legend). The consensus secondary structure, predicted based on the multiple sequence alignment, is encoded in dot bracket format (see first row) and also shown in Fig. 3A inset.

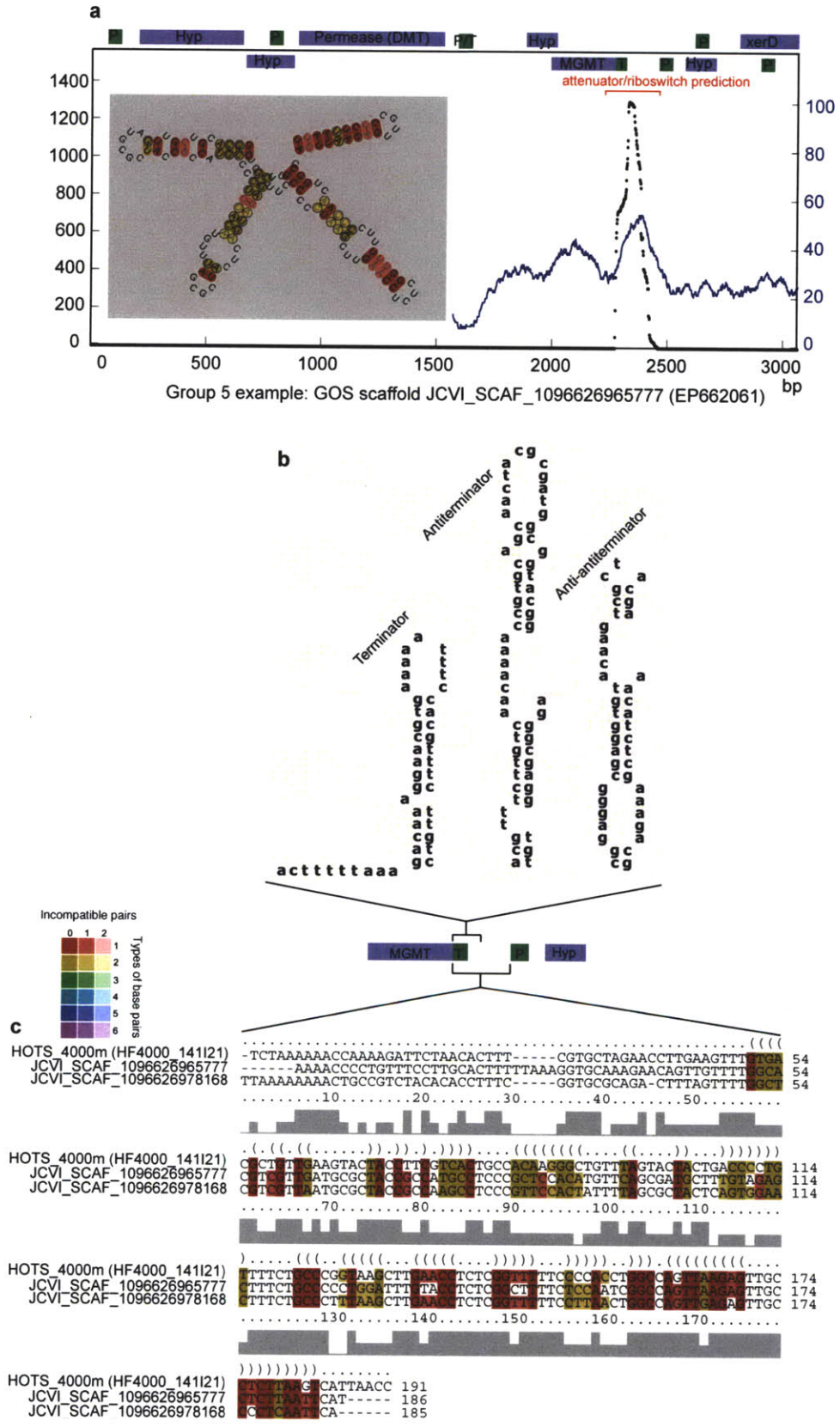

a
Hyp
Permease (DMT)
Hyp
Hyp
xerD
MGMT
Hyp
attenuator/riboswitch prediction
1400
1200
1000
800
600
400
200
0
100
80
60
40
20
0
0
500
1000
1500
2000
2500
3000
bp
Group 5 example: GOS scaffold JCVI_SCAF_1096626965777 (EP662061)
b
Antiterminator
Anti-antiterminator
Terminator
actttttaaa
MGMT
Hyp
Incompatible pairs
0 1 2
1
2
3
4
5
6
Types of base pairs
c
HOTS_4000m (HF4000_141I21)
JCVI_SCAF_1096626965777
JCVI_SCAF_1096626978168
54
54
54
10
20
30
40
50
114
114
114
70
80
90
100
110
174
174
174
130
140
150
160
170
191
186
185

Figure S6 (previous page). Genomic context and secondary structure prediction of Group 5 psRNA. **(a)** The Group 5 psRNA sequences mapped onto the reference metagenomic fragment (JCVI_SCAF_10 96626965777), with sequence coverage (black dots, left axis) and reference GC content (blue dots, right axis) shown. Gene annotations are indicated along the top of each panel (upper and lower lines represent forward and reverse strands; P and T represent promoter and terminator, respectively). The consensus secondary structure shown in the inset was predicted based on the multiple sequence alignment shown in panel c. **(b)** Attenuator-like structures including terminator, antiterminator and anti-antiterminator in the 5'-UTR of the 6-O-MGMT gene were predicted using RibEx (Abreu-Goodger & Merino, 2005). **(c)** The alignment of Group 5 psRNAs from three *Pelagibacter*-like genomic fragments, including one from 4000m deep ocean, and two from surface open ocean. The alignment is color-coded according to the different types of base pairs and the amount of compensatory and incompatible base changes in the corresponding alignment columns (see color legend above).

CHAPTER SIX

Summary and future directions

Chapter 6: Summary and future directions

Summary

Metagenomic research has paved the way for a comprehensive understanding of microbial gene parts list, but our understanding of the expression, regulation, function, and ecological relevance of these genes has proceeded more slowly. This thesis work has provided methodological foundation for obtaining and analyzing metatranscriptomic data from natural microbial assemblages. Application of metatranscriptomics in both survey and experimental settings has further contributed towards a better understanding of microbial gene expression and regulation in natural settings, as well as the environmental factors (biotic and abiotic) that influence microbial assemblage dynamics in the open ocean. The main findings from this body of work are summarized below.

Chapter 2. Methodology development, validation, and pilot study of microbial metatranscriptomics

1. Microbial community transcriptomes can be profiled (for abundant taxa, and highly expressed genens), and interpreted in the context of taxonomic structure, genomic composition, and ambient environmental conditions.

2. Metatranscriptomic data are characterized by a wealth of novel transcripts that are often of unknown function or phylogenetic origin, and that have not been detected or only rarely detected in publicly available DNA databases.

Chapter 3. Integrated metatranscriptomic and metagenomic analyses of 4 bacterioplankton samples in the water column

1. Based on functional assignments, metatranscriptomic samples cluster to the exclusion of corresponding metagenomic data sets, likely resulting from the active expression of house-keeping genes. Clustering among metatranscriptomic data sets however, correlates with the spatial and temporal relatedness of samples.

2. Habitat-specific metabolic processes are discernible at the transcriptional level, and can

sometimes be attributed to specific taxa. Examples include *Roseobacter*-relatives involved in aerobic anoxygenic phototrophy at 75-m depth, and the unexpected contribution of low abundance *Crenarchaea* to ammonia oxidation at 125-m depth.

3. Taxonomic representation can significantly differ between cDNA and corresponding DNA samples, highlighting the decoupling of abundance and activity. Numerically less abundant microorganisms may nevertheless contribute actively to ecologically relevant processes.

4. Genome-centric analyses of representative taxa including *Pelagibacter* and *Prochlorococcus* show transcriptional signals consistent with known physiology or protein expression profiles in the laboratory.

Chapter 4. A case study for understanding how an environmental driver, in this case, nutrient loading via deep water mixing, can affect microbial transcriptional profiles.

1. Some taxa that are present in low abundance in normal conditions may respond quickly to environmental perturbation, by displaying chemotactic behavior and active cell growth.

2. Dynamics of phage-host interactions appeared to have been altered by nutrient loading from the deep seawater. Specifically, captured cyanophage DNA and cDNA profiles resembled possible transition from phage pseudolysogeny to active lysis. This hypothesis, if validated, has significant ecological relevance given the critical roles of phage activities in biogeochemical cycling and genetic diversity.

3. Microbial responses observed at the transcriptional level on a shorter time scale (hours), provide insights into mechanisms that lead to the community dynamics observed on a longer time scale (days to weeks). An example here is that *Prochlorococcus* cells, frequently observed to be outcompeted by larger phytoplankton during deep mixing, displayed elevated gene expression for carbon fixation and photosynthesis, as well as higher cell density, relative to the control, during the first 27 hours. This observation, in the context of the community transcriptome, suggested that previously reported phytoplankton shift from *Prochlorococcus* to larger cells might not be due to decrease in *Prochlorococcus* cellular fitness but more likely caused by higher phage-induced mortality and possibly grazing rate.

Chapter 5. The unexpected discovery of highly expressed small noncoding RNA transcripts, and the characterization of their genomic context, sequence variability, and structural properties.

1. With metatranscriptomic analysis it is now feasible to study naturally occurring noncoding RNA elements, including riboswitches, and cis- and trans-regulators, that are highly expressed in natural microbial populations and in many cases appear to be derived from as-yet uncharacterized microorganisms.

2. The extraordinary abundance of some of the identified small RNAs suggests their potential functional significance, which can be investigated with respect to their genomic context, but remain to be elucidated in model systems.

3. The universal presence of highly expressed small RNAs in metatranscriptomic data sets suggests that small RNA regulation is the rule rather than the exception in microbial gene regulation in ocean waters.

Future directions

This thesis work has advanced our knowledge on the composition and dynamics of microbial community transcriptomes *in situ*. At the same time, this work has also raised questions for future investigation.

First, how do metatranscriptomic data translate to the rates of specific geochemical processes? Being able to answer this question is a long-term goal but nonetheless a critical one, for the following reasons. Researchers have been striving to understand how transcript abundance relates to cognate protein levels, and metabolic rates in model systems. Given what is already known, the interplay among these measurements at the community level will undoubtedly be orders of magnitude more complex. But advances in understanding this interplay would move us forward towards using community transcriptome profiles not only to generate new hypotheses (as we are doing now), but also to quantitatively assess specific geochemical processes mediated

by the microorganisms. Studies like the one led by Don Canfield (Canfield et al., 2010) where the authors combined molecular techniques and high resolution process rate measurements are essential steps towards this goal.

In Chapter 4, we were able to monitor the composition of microbial community transcriptomes in a microcosm experiment over time for 27 hours. The results are gratifying in that the temporal dynamics suggests how different taxa interact and evolve over time, suggesting possible mechanisms that lead to bulk-level changes (for instance, community structure shifts, primary and bacterial production, etc.). Along the same lines, it would be helpful to perform time-series surveys on community transcriptomes, which can expand our knowledge of snapshots of microbial gene expression to a more realistic view of the gene expression dynamics. The DeLong lab has initiated the collection of RNA samples at the Hawaii Ocean Time-series (HOT) station ALOHA, on a monthly basis, but subjecting all these RNA samples to deep sequencing is currently impractical and too costly. In particular, the high content of transcripts with house-keeping functions (e.g., rRNAs, tRNAs, ribosomal protein RNAs, etct.) results in the relatively low sequencing coverage for genes involved in habitat-specific functions. Removal of rRNAs (Stewart et al., 2010) and cDNA library normalization prior to sequencing (Rodrigue et al., 2009) is one potential solution. Alternatively, one can apply custom-designed microarrays ((Rich, Pham, Eppley, Shi & DeLong, 2010); Appendix B) to screen RNA samples in a low cost and high-throughput fashion, and consequently to identify those samples with interesting or unique signals for further deep sequencing (at a higher coverage).

An unexpected but exciting finding from metatranscriptomic studies is the wealth of novel noncoding small RNAs (sRNAs), which, judging from their abundance and diversity (some clearly are derived from phages), must play important roles in nature. We gained some insight into the potential targets of these sRNAs by using computational methods based on thermodynamic pairing energies and known sRNA-mRNA hybrids, but knowledge of their biochemical functions is key to grasping the essential significance of such highly expressed sRNA elements. *Prochlorococcus* and *Pelagibacter*, two model organisms in culture that are also abundant in nature, provide useful platforms for such sRNA-centered studies (Meyer et al.,

2009; Steglich et al., 2008). The real challenge however, is that many novel sRNAs appear to be derived from as-yet-uncultivated microorganisms, raising the need of studying these sRNAs *in vitro* (Meyer, Roth, Chervin, Garcia & Breaker, 2008) or in a heterologous host system (Said et al., 2009). I planned an experiment (Figure 1) to screen for sRNA target genes, which takes advantage of controllable heterologous expression of sRNA genes and the large archive of fosmid clones. Due to time limitations, I was not able to complete these experiments, but they are certainly worth pursuing in the future.

Finally, some interesting but unclear signals have emerged from the studies presented in this thesis, and may be worth following up in the future. For example, metatranscriptomic sequences are consistently found to bear higher GC content than the corresponding metagenomic sequences. This could be caused by higher representation of high GC content genomes in the metatranscriptomic data or by preferred active expression of high GC content ORFs. It seems that the latter is more likely, based on a closer inspection of expressed ORFs from *Pelagibacter* genome (characterized by low GC-content). The top 10 most highly expressed ORFs on the *Pelagibacter* genome fall above the 90th percentile in GC content among all ORFs (Data not shown). Such correlation was proposed for mammalian chromosomes (Konu & Li, 2002; Semon, Mouchiroud & Duret, 2005). Other models of microbial gene expression include: 1) expression variations of genes are proportional to their express levels (Ueda et al., 2004); 2) Expression levels depend to mRNA structure (specifically, the 5'-UTR of mRNA) (Kudla, Murray, Tollervey & Plotkin, 2009); and 3) Gene expression levels influence amino acid usage (Schaber et al., 2005). Metatranscriptomics can serve as a superior platform for testing the generality of these hypotheses in the future.

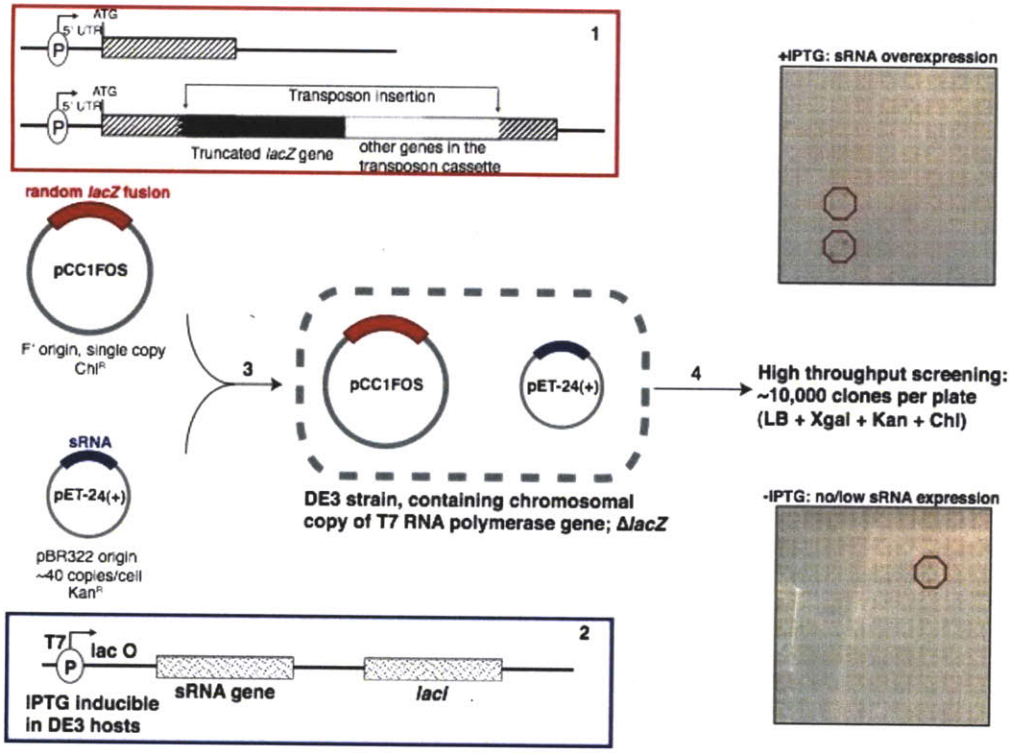


Figure 1. Schematic illustration of sRNA target gene screening experiment. Step 1 is the construction of reporter gene fusion by transposon insertion of truncated *lacZ* gene to fosmid clones. Step 2 is the construction of sRNA plasmid whose expression is IPTG-inducible. Step 3 involves the transformation of both constructs to *E. coli* host cells. Step 4 is the macroarray screening based on blue-white phenotype.

Bibliography

Abbott, G. (1999). Proteomics, transcriptomics; what's in a name. *Nature, 202*, 715-716.

Abreu-Goodger, C. & Merino, E. (2005). Ribex: A web server for locating riboswitches and other conserved bacterial regulatory elements. *Nucleic Acids Research, 33*, W690-W692.

Acinas, S. G., Anton, J., & Rodriguez-Valera, F. (1999). Diversity of free-living and attached bacteria in offshore western mediterranean waters as depicted by analysis of genes encoding 16S rrna. *Applied and Environmental Microbiology, 65*(2), 514-522.

Acinas, S. G., Klepac-Ceraj, V., Hunt, D. E., Pharino, C., Ceraj, I., Distel, D. L., et al. (2004). Fine-Scale phylogenetic architecture of a complex bacterial community. *Nature, 430*(6999), 551-554.

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., & Lipman, D. J. (1990). Basic local alignment search tool. *Journal of Molecular Biology, 215*(3), 403-410.

Andersson, A. F., Lundgren, M., Eriksson, S., Rosenlund, M., Bernander, R., & Nilsson, P. (2006). Global analysis of mrna stability in the archaeon sulfolobus. *Genome Biology, 7*(10), R99.

Angly, F. E., Felts, B., Breitbart, M., Salamon, P., Edwards, R. A., Carlson, C., et al. (2006). The marine viromes of four oceanic regions. *Plos Biology, 4*(11), e368.

Arrigo, K. R. (2005). Marine microorganisms and global nutrient cycles. *Nature, 437*(7057), 349-355.

Aumont, O., Maier-Reimer, E., Blain, S., & Monfray, P. (2003). An ecosystem model of the global ocean including fe, si, P colimitations. *Global Biogeochemical Cycles, 17*(2), 1060.

Axmann, I. M., Kensche, P., Vogel, J., Kohl, S., Herzel, H., & Hess, W. R. (2005). Identification of cyanobacterial non-coding rnas by comparative genome analysis. *Genome Biology, 6*(9), R73.

Bailly, J., Fraissinet-Tachet, L., Verner, M. C., Debaud, J. C., Lemaire, M., Wésolowski-Louvel, M., et al. (2007). Soil eukaryotic functional diversity, a metatranscriptomic approach. *Isme Journal, 1*(7), 632-42.

Barrick, J. E., Sudarsan, N., Weinberg, Z., Ruzzo, W. L., & Breaker, R. R. (2005). 6S RNA is a widespread regulator of eubacterial RNA polymerase that resembles an open promoter. *Rna-A Publication of the Rna Society, 11*(5), 774-784.

Berube, P. M., Samudrala, R., & Stahl, D. A. (2007). Transcription of all amoc copies is associated with recovery of nitrosomonas europaea from ammonia starvation. *Journal of Bacteriology, 189*(11), 3935-3944.

Béjà, O., Aravind, L., Koonin, E. V., Suzuki, M. T., Hadd, A., Nguyen, L. P., et al. (2000). Bacterial rhodopsin: Evidence for a new type of phototrophy in the sea. *Science, 289*(5486), 1902-1906.

Béjà, O., Spudich, E. N., Spudich, J. L., Leclerc, M., & DeLong, E. F. (2001). Proteorhodopsin phototrophy in the ocean. *Nature, 411*(6839), 786-789.

Béjà, O., Suzuki, M. T., Heidelberg, J. F., Nelson, W. C., Preston, C. M., Hamada, T., et al. (2002). Unsuspected diversity among marine aerobic anoxygenic phototrophs. *Nature*, *415*(6872), 630-633.

Bolstad, B. M., Irizarry, R. A., Astrand, M., & Speed, T. P. (2003). A comparison of normalization methods for high density oligonucleotide array data based on variance and bias. *Bioinformatics*, *19*(2), 185-193.

Boucher, Y., Douady, C. J., Papke, R. T., Walsh, D. A., Boudreau, M. E. R., Nesbo, C. L., et al. (2003). Lateral gene transfer and the origins of prokaryotic groups. *Annual Review of Genetics*, *37*, 283-328.

Braddock, J. F., Ruth, M. L., Catterall, P. H., Walworth, J. L., & McCarthy, K. A. (1997). Enhancement and inhibition of microbial activity in hydrocarbon-contaminated arctic soils: Implications for nutrient-amended bioremediation. *Environ. Sci. Technol*, *31*(7), 2078-2084.

Brantl, S. (2004). Bacterial gene regulation: From transcription attenuation to riboswitches and ribozymes. *Trends in Microbiology*, *12*(11), 473-475.

Brazelton, W. J., Ludwig, K. A., Sogin, M. L., Andreishcheva, E. N., Kelley, D. S., Shen, C. C., et al. (2010). Archaea and bacteria with surprising microdiversity show shifts in dominance over 1,000-year time scales in hydrothermal chimneys. *Proceedings of the National Academy of Sciences of the United States of America*, *107*(4), 1612-1617.

Breaker, R. R. (2008). Complex riboswitches. *Science (New York, N.Y.)*, *319*(5871), 1795-1797.

Brown, J. M. & Hewson, I. (2010). Ecophysiology of a common unannotated gene transcript in surface water microbial assemblages of the oligotrophic open ocean. *AQUAT MICROB ECOL*, *60*(3), 289-297.

Bruttin, A. & Brüssow, H. (1996). Site-Specific spontaneous deletions in three genome regions of a temperate *streptococcus thermophilus* phage. *Virology*, *219*(1), 96-104.

Brzezinski, M. A. (1988). Vertical-Distribution of ammonium in stratified oligotrophic waters. *Limnology and Oceanography*, *33*(5), 1176-1182.

Campbell, L., Liu, H., Nolla, H. A., & Vaulot, D. (1997). Annual variability of phytoplankton and bacteria in the subtropical north pacific ocean at station ALOHA during the 1991-1994 ENSO event. *Deep Sea Research Part I: Oceanographic Research Papers*, *44*(2), 167-192.

Campbell, L., Nolla, H. A., & Vaulot, D. (1994). The importance of prochlorococcus to community structure in the central north pacific ocean. *Limnology and Oceanography*, *39*(4), 954-961.

Canfield, D. E., Stewart, F. J., Thamdrup, B., De Brabandere, L., Dalsgaard, T., Delong, E. F., et al. (2010). A cryptic sulfur cycle in oxygen-minimum-zone waters off the chilean coast. *Science (New York, N.Y.)*, *1375-1378.*

Cappello, S., Caruso, G., Zampino, D., Monticelli, L. S., Maimone, G., Denaro, R., et al. (2007). Microbial community dynamics during assays of harbour oil spill bioremediation: A microscale simulation study. *Journal of Applied Microbiology*, *102*(1), 184-194.

Carlson, C. A., Ducklow, H. W., Hansell, D. A., & Smith Jr, W. O. (1998). Organic carbon partitioning during spring phytoplankton blooms in the ross sea polynya and the sargasso sea. *Limnology and Oceanography*, *43*(3), 375-386.

Carlson, C. A., Giovannoni, S. J., Hansell, D. A., Goldberg, S. J., Parsons, R., & Vergin, K. (2004). Interactions among dissolved organic carbon, microbial processes, and community structure in the mesopelagic zone of the northwestern sargasso sea. *Limnology and Oceanography*, *49*(4), 1073-1083.

Carlson, C. A., Stephen, J. G., Dennis, A. H., Stuart, J. G., Rachel, P., Mark, P. O., et al. (2002). Effect of nutrient amendments on bacterioplankton production, community structure, and DOC utilization in the northwestern sargasso sea. *Aquat. Microb. Ecol*, *30*, 19-36.

Casciotti, K. L. & Ward, B. B. (2001). Dissimilatory nitrite reductase genes from autotrophic ammonia-oxidizing bacteria. *Applied and Environmental Microbiology*, *67*(5), 2213-2221.

Church, M. J., Wai, B., Karl, D. M., & DeLong, E. F. (2010). Abundances of crenarchaeal amoa genes and transcripts in the pacific ocean. *Environmental Microbiology*, *12*(3), 679-688.

Coleman, M. L. & Chisholm, S. W. (2007). Code and context: Prochlorococcus as a model for cross-scale biology. *Trends in Microbiology*, *15*(9), 398-407.

Coleman, M. L., Sullivan, M. B., Martiny, A. C., Steglich, C., Barry, K., DeLong, E. F., et al. (2006). Genomic islands and the ecology and evolution of prochlorococcus. *Science*, *311*(5768), 1768-1770.

Corbin, R. W., Paliy, O., Yang, F., Shabanowitz, J., Platt, M., Lyons, C. E., et al. (2003). Toward a protein profile of escherichia coli: Comparison to its transcription profile. *Proc. Natl Acad. Sci. USA*, *100*(16), 9232-9237.

Croft, L., Lercher, M., Gagen, M., & Mattick, J. (2003). Is prokaryotic complexity limited by accelerated growth in regulatory overhead? *Genome Biology*, *5*(1), P2.

Cuadros-Orellana, S., Martin-Cuadrado, A. B., Legault, B., D'Auria, G., Zhaxybayeva, O., Papke, R. T., et al. (2007). Genomic plasticity in prokaryotes: The case of the square haloarchaeon. *Isme Journal*, *1*(3), 235-245.

Curtis, T. P., Sloan, W. T., & Scannell, J. W. (2002). Estimating prokaryotic diversity and its limits. *Proc. Natl Acad. Sci. USA*, *99*(16), 10494-10499.

Dafforn, A., Chen, P., Deng, G., Herrler, M., Iglehart, D., Koritala, S., et al. (2004). Linear mrna amplification from as little as 5 ng total RNA for global gene expression analysis. *Biotechniques*, *37*(5), 854-857.

DeLong, E. F. & Béjà, O. (2010). The light-driven proton pump proteorhodopsin enhances bacterial survival during tough times. *Plos Biology*, *8*(4), e1000359.

DeLong, E. F. & Karl, D. M. (2005). Genomic perspectives in microbial oceanography. *Nature*, *437*(7057), 336-342.

DeLong, E. F., Preston, C. M., Mincer, T., Rich, V., Hallam, S. J., Frigaard, N. U., et al. (2006). Community genomics among stratified microbial assemblages in the ocean's interior. *Science*, *311*(5760), 496-503.

DeSantis, T. Z., Hugenholtz, P., Larsen, N., Rojas, M., Brodie, E. L., Keller, K., et al. (2006). Greengenes, a chimera-checked 16S rrna gene database and workbench compatible with ARB. *Applied and Environmental Microbiology*, *72*(7), 5069-5072.

Dinsdale, E. A., Edwards, R. A., Hall, D., Angly, F., Breitbart, M., Brulc, J. M., et al. (2008). Functional metagenomic profiling of nine biomes. *Nature*, *452*(7187), 629-632.

Dore, J. E. & Karl, D. M. (1996). Nitrite distributions and dynamics at station ALOHA. *Deep Sea Research Part II: Topical Studies in Oceanography*, *43*(2-3), 385-402.

Dore, J. E., Letelier, R. M., Church, M. J., Lukas, R., & Karl, D. M. (2008). Summer phytoplankton blooms in the oligotrophic north pacific subtropical gyre: Historical perspective and recent observations. *Progress in Oceanography*, *76*(1), 2-38.

Duehring, U., Axmann, I. M., Hess, W. R., & Wilde, A. (2006). An internal antisense RNA regulates expression of the photosynthesis gene *isia. Proc. Natl Acad. Sci. USA*, *103*(18), 7054-7058.

Dufresne, A., Salanoubat, M., Partensky, F., Artiguenave, F., Axmann, I. M., Barbe, V., et al. (2003). Genome sequence of the cyanobacterium prochlorococcus marinus SS120, a nearly minimal oxyphototrophic genome. *Proc. Natl Acad. Sci. USA*, *100*(17), 10020-10025.

Eddy, S. R. & Durbin, R. (1994). RNA sequence-analysis using covariance-models. *Nucleic Acids Research*, *22*(11), 2079-2088.

Eddy, S. (2007). INFERNAL user's guide, version 0.72. *INFERNAL User's Guide, Version 0.72.*

Edgar, R. C. (2004). MUSCLE: Multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research*, *32*(5), 1792-1797.

Eiler, A., Hayakawa, D. H., Church, M. J., Karl, D. M., & Rappé, M. S. (2009). Dynamics of the SAR11 bacterioplankton lineage in relation to environmental conditions in the oligotrophic north pacific subtropical gyre. *Environmental Microbiology*, *11*(9), 2291-2300.

Eilers, H., Pernthaler, J., Glöckner, F. O., & Amann, R. (2000). Culturability and in situ abundance of pelagic bacteria from the north sea. *Applied and Environmental Microbiology*, *66*(7), 3044-3051.

Ernst, F. D., Bereswill, S., Waidner, B., Stoof, J., Mäder, U., Kusters, J. G., et al. (2005). Transcriptional profiling of helicobacter pylori fur- and iron-regulated gene expression. *Microbiology (Reading, England)*, *151*(Pt 2), 533-546.

Eymann, C., Homuth, G., Scharf, C., & Hecker, M. (2002). Bacillus subtilis functional genomics: Global characterization of the stringent response by proteome and transcriptome analysis. *Journal of Bacteriology*, *184*(9), 2500–2520.

Feldman, A. L., Costouros, N. G., Wang, E., Qian, M., Marincola, F. M., Alexander, H. R., et al. (2002). Advantages of mrna amplification for microarrfriay analysis. *Biotechniques*, *33*(4), 906-914.

Francis, C. A., Co, E. M., & Tebo, B. M. (2001). Enzymatic manganese(II) oxidation by a marine alpha-proteobacterium. *Applied and Environmental Microbiology*, *67*(9), 4024-

4029.

Frias-Lopez, J., Shi, Y., Tyson, G. W., Coleman, M. L., Schuster, S. C., Chisholm, S. W., et al. (2008). Microbial community gene expression in ocean surface waters. *Proceedings of the National Academy of Sciences of the United States of America*, *105*(10), 3805-3810.

Frigaard, N. U., Martinez, A., Mincer, T. J., & DeLong, E. F. (2006). Proteorhodopsin lateral gene transfer between marine planktonic bacteria and archaea. *Nature*, *439*(7078), 847-850.

Fuhrman, J. A. & Azam, F. (1980). Bacterioplankton secondary production estimates for coastal waters of british columbia, antarctica, and california. *Applied and Environmental Microbiology*, *39*(6), 1085-1095.

Gifford, S. M., Sharma, S., Rinta-Kanto, J. M., & Moran, M. A. (2010). Quantitative analysis of a deeply sequenced marine microbial metatranscriptome. *The ISME Journal*, doi:10.1038/ismej.2010.141.

Gilbert, J. A., Field, D., Huang, Y., Edwards, R., Li, W., Gilna, P., et al. (2008). Detection of large numbers of novel sequences in the metatranscriptomes of complex marine microbial communities. *Plos ONE*, *3*(8), e3042.

Gill, S. R., Pop, M., Deboy, R. T., Eckburg, P. B., Turnbaugh, P. J., Samuel, B. S., et al. (2006). Metagenomic analysis of the human distal gut microbiome. *Science (New York, N.Y.)*, *312*(5778), 1355-1359.

Giovannoni, S. J. & Stingl, U. (2005). Molecular diversity and ecology of microbial plankton. *Nature*, *437*(7057), 343-348.

Giovannoni, S. J., Bibbs, L., Cho, J. C., Stapels, M. D., Desiderio, R., Vergin, K. L., et al. (2005a). Proteorhodopsin in the ubiquitous marine bacterium SAR11. *Nature*, *438*(7064), 82-85.

Giovannoni, S. J., Tripp, H. J., Givan, S., Podar, M., Vergin, K. L., Baptista, D., et al. (2005b). Genome streamlining in a cosmopolitan oceanic bacterium. *Science*, *309*(5738), 1242-1245.

González, J. M., Fernández-Gómez, B., Fernàndez-Guerra, A., Gómez-Consarnau, L., Sánchez, O., Coll-Lladó, M., et al. (2008). Genome analysis of the proteorhodopsin-containing marine bacterium polaribacter sp. MED152 (flavobacteria). *Proceedings of the National Academy of Sciences of the United States of America*, *105*(25), 8724-8729.

Gottesman, S. (2002). Stealth regulation: Biological circuits with small RNA switches. *Genes & Development*, *16*(22), 2829-2842.

Gottesman, S. (2004). The small RNA regulators of escherichia coli: Roles and mechanisms. *Annual Review of Microbiology*, *58*, 303-328.

Gómez-Consarnau, L., Akram, N., Lindell, K., Pedersen, A., Neutze, R., Milton, D. L., et al. (2010). Proteorhodopsin phototrophy promotes survival of marine bacteria during starvation. *Plos Biology*, *8*(4), e1000358.

Gómez-Consarnau, L., Gonzalez, J. M., Coll-Llado, M., Gourdon, P., Pascher, T., Neutze, R., et al. (2007). Light stimulates growth of proteorhodopsin-containing marine flavobacteria.

Nature, *445*(7124), 210-213.

Griffiths-Jones, S., Moxon, S., Marshall, M., Khanna, A., Eddy, S. R., & Bateman, A. (2005). Rfam: Annotating non-coding rnas in complete genomes. *Nucleic Acids Research*, *33*(Database issue), 121-124.

Guell, M., van Noort, V., Yus, E., Chen, W. -H., Leigh-Bell, J., Michalodimitrakis, K., et al. (2009). Transcriptome complexity in a genome-reduced bacterium. *Science*, *326*(5957), 1268-1271.

Hallam, S. J., Mincer, T. J., Schleper, C., Preston, C. M., Roberts, K., Richardson, P. M., et al. (2006). Pathways of carbon assimilation and ammonia oxidation suggested by environmental genomic analyses of marine crenarchaeota. *Plos Biology*, *4*(4), 520-536.

Hansell, D. A. & Carlson, C. A. (2001). Biogeochemistry of total organic carbon and nitrogen in the sargasso sea: Control by convective overturn. *Deep-Sea Research Part Ii-Topical Studies in Oceanography*, *48*(8-9), 1649-1667.

He, S. M., Kunin, V., Haynes, M., Martin, H. G., Ivanova, N., Rohwer, F., et al. (2010a). Metatranscriptomic array analysis of 'candidatus accumulibacter phosphatis'-enriched enhanced biological phosphorus removal sludge. *ENVIRON MICROBIOL*, *12*(5), 1205-1217.

He, S., Wurtzel, O., Singh, K., Froula, J. L., Yilmaz, S., Tringe, S. G., et al. (2010b). Validation of two ribosomal RNA removal methods for microbial metatranscriptomics. *Nature Methods*, *7*(10), 807-812.

Hershberg, R., Altuvia, S., & Margalit, H. (2003). A survey of small rna-encoding genes in *escherichia coli. Nucleic Acids Research*, *31*(7), 1813-1820.

Hewson, I., Poretsky, R. S., Beinart, R. A., White, A. E., Shi, T., Bench, S. R., et al. (2009a). In situ transcriptomic analysis of the globally important keystone n-2-fixing taxon crocosphaera watsonii. *Isme Journal*, *3*(5), 618-631.

Hewson, I., Poretsky, R. S., Dyhrman, S. T., Zielinski, B., White, A. E., Tripp, H. J., et al. (2009b). Microbial community gene expression within colonies of the diazotroph, trichodesmium, from the southwest pacific ocean. *Isme Journal*, *3*(11), 1286-1300.

Hewson, I., Rachel, S. P., Tripp, H. J., Joseph, P. M., & Jonathan, P. Z. (2010). Spatial patterns and light-driven variation of microbial population gene expression in surface waters of the oligotrophic open ocean. *Environmental Microbiology*, *12*(7), 1940-1956.

Hofacker, I. L. (2003). Vienna RNA secondary structure server. *Nucleic Acids Research*, *31*(13), 3429-3431.

Hofacker, I. L., Fekete, M., & Stadler, P. F. (2002). Secondary structure prediction for aligned RNA sequences. *Journal of Molecular Biology*, *319*(5), 1059-1066.

Hollibaugh, J. T., Gifford, S., Sharma, S., Bano, N., & Moran, M. A. (2010). Metatranscriptomic analysis of ammonia-oxidizing organisms in an estuarine bacterioplankton assemblage. *The ISME Journal*, doi:10.1038/ismej.2010.172.

Holste, D., Weiss, O., Grosse, I., & Herzel, H. (2000). Are noncoding sequences of *rickettsia prowazekii* remnants of "neutralized" genes? *Journal of Molecular Evolution*, *51*(4), 353-

362.

Holtzendorff, J., Partensky, F., Jacquet, S., Bruyant, F., Marie, D., Garczarek, L., et al. (2001). Diel expression of cell cycle-related genes in synchronized cultures of prochlorococcus sp strain PCC 9511. *Journal of Bacteriology*, *183*(3), 915-920.

Hunt, D. E., David, L. A., Gevers, D., Preheim, S. P., Alm, E. J., & Polz, M. F. (2008). Resource partitioning and sympatric differentiation among closely related bacterioplankton. *Science (New York, N.Y.)*, *320*(5879), 1081-1085.

Huse, S., Huber, J., Morrison, H., Sogin, M., & Welch, D. (2007). Accuracy and quality of massively parallel DNA pyrosequencing. *Genome Biology*, *8*(7), R143.

Huson, D. H., Auch, A. F., Qi, J., & Schuster, S. C. (2007). MEGAN analysis of metagenomic data. *Genome Research*, *17*(3), 377-386.

Ivars-Martinez, E., Martin-Cuadrado, A. B., D'Auria, G., Mira, A., Ferriera, S., Johnson, J., et al. (2008). Comparative genomics of two ecotypes of the marine planktonic copiotroph alteromonas macleodii suggests alternative lifestyles associated with different kinds of particulate organic matter. *The ISME Journal*, *2*(12), 1194-1212.

Ivars-Martínez, E., D'Auria, G., Rodríguez-Valera, F., Sánchez-Porro, C., Ventosa, A., Joint, I., et al. (2008). Biogeography of the ubiquitous marine bacterium alteromonas macleodii determined by multilocus sequence analysis. *Molecular Ecology*, *17*, 4092-4106.

Jason, S. & Ewan, B. (2000). The bioperl project: Motivation and usage. *SIGBIO Newsl.*, *20*(2), 13-14.

Johnson, Z. I., Zinser, E. R., Coe, A., McNulty, N. P., Woodward, E. M. S., & Chisholm, S. W. (2006). Niche partitioning among prochlorococcus ecotypes along ocean-scale environmental gradients. *Science*, *311*(5768), 1737-1740.

Kanehisa, M. & Goto, S. (2000). KEGG: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Research*, *28*(1), 27-30.

Karl, D. M. & Letelier, R. M. (2008). Nitrogen fixation-enhanced carbon sequestration in low nitrate, low chlorophyll seascapes. *Marine Ecology-Progress Series*, *364*, 257-268.

Karl, D. M. & Lukas, R. (1996). The hawaii ocean time-series (HOT) program: Background, rationale and field implementation. *Deep Sea Research Part II: Topical Studies in Oceanography*, *43*(2-3), 129-156.

Karl, D. M. (2002). Nutrient dynamics in the deep blue sea. *Trends in Microbiology*, *10*(9), 410-418.

Karl, D. M. (2007). Microbial oceanography: Paradigms, processes and promise. *Nature Reviews. Microbiology*, *5*(10), 759-769.

Karl, D. M., Beversdorf, L., Björkman, K. M., Church, M. J., Martinez, A., & Delong, E. F. (2008). Aerobic production of methane in the sea. *Nature Geoscience*, *1*(7), 473-478.

Kawano, M., Reynolds, A. A., Miranda-Rios, J., & Storz, G. (2005). Detection of 5'- and 3'-utr-derived small rnas and cis-encoded antisense rnas in escherichia coli. *Nucleic Acids Research*, *33*(3), 1040-1050.

Kettler, G. C., Martiny, A. C., Huang, K., Zucker, J., Coleman, M. L., Rodrigue, S., et al. (2007).

Patterns and implications of gene gain and loss in the evolution of prochlorococcus. *Plos Genetics*, *3*(12), e231.

Kolber, Z. S., Plumley, F. G., Lang, A. S., Beatty, J. T., Blankenship, R. E., VanDover, C. L., et al. (2001). Contribution of aerobic photoheterotrophic bacteria to the carbon cycle in the ocean. *Science*, *292*(5526), 2492-2495.

Konneke, M., Bernhard, A. E., de la Torre, J. R., Walker, C. B., Waterbury, J. B., & Stahl, D. A. (2005). Isolation of an autotrophic ammonia-oxidizing marine archaeon. *Nature*, *437*(7058), 543-546.

Konstantinidis, K. T. & DeLong, E. F. (2008). Genomic patterns of recombination, clonal divergence and environment in marine microbial populations. *The ISME Journal*, *2*, 1052-1065.

Konu, O. & Li, M. D. (2002). Correlations between mrna expression levels and GC contents of coding and untranslated regions of genes in rodents. *Journal of Molecular Evolution*, *54*(1), 35-41.

Kort, R., Keijser, B. J., Caspers, M. P., Schuren, F. H., & Montijn, R. (2008). Transcriptional activity around bacterial cell death reveals molecular biomarkers for cell viability. *BMC Genomics*, *9*, 590.

Kudla, G., Murray, A. W., Tollervey, D., & Plotkin, J. B. (2009). Coding-Sequence determinants of gene expression in escherichia coli. *Science (New York, N.Y.)*, *324*(5924), 255-258.

Lami, R., Cottrell, M. T., Campbell, B. J., & Kirchman, D. L. (2009). Light-Dependent growth and proteorhodopsin expression by flavobacteria and SAR11 in experiments with delaware coastal waters. *Environmental Microbiology*, *11*(12), 3201-3209.

Leininger, S., Urich, T., Schloter, M., Schwark, L., Qi, J., Nicol, G. W., et al. (2006). Archaea predominate among ammonia-oxidizing prokaryotes in soils. *Nature*, *442*(7104), 806-809.

Lenz, D. H., Mok, K. C., Lilley, B. N., Kulkarni, R. V., Wingreen, N. S., & Bassler, B. L. (2004). The small RNA chaperone hfq and multiple small rnas control quorum sensing in vibrio harveyi and vibrio cholerae. *Cell*, *118*(1), 69-82.

Letunic, I. & Bork, P. (2007). Interactive tree of life (itol): An online tool for phylogenetic tree display and annotation. *Bioinformatics (Oxford, England)*, *23*(1), 127-128.

Li, W. Z. & Godzik, A. (2006). Cd-Hit: A fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics*, *22*(13), 1658-1659.

Lin, S. J., Zhang, H. A., Zhuang, Y. Y., Tran, B., & Gill, J. (2010). Spliced leader-based metatranscriptomic analyses lead to recognition of hidden genomic features in dinoflagellates. *P NATL ACAD SCI USA*, *107*(46), 20033-20038.

Lindell, D. & Post, A. F. (1995). Ultraphytoplankton succession is triggered by deep winter mixing in the gulf-of-aqaba (eilat), red-sea. *Limnology and Oceanography*, *40*(6), 1130-1141.

Lindell, D., Erdner, D., Marie, D., Prasil, O., Koblizek, M., Le Gall, F., et al. (2002). Nitrogen stress response of prochlorococcus strain PCC 9511 (oxyphotobacteria) involves

contrasting regulation of ntca and amt1. *Journal of Phycology*, *38*(6), 1113-1124.

Lindell, D., Jaffe, J. D., Coleman, M. L., Futschik, M. E., Axmann, I. M., Rector, T., et al. (2007). Genome-Wide expression dynamics of a marine virus and host reveal features of co-evolution. *Nature*, *449*(7158), 83-86.

Lindell, D., Sullivan, M. B., Johnson, Z. I., Tolonen, A. C., Rohwer, F., & Chisholm, S. W. (2004). Transfer of photosynthesis genes to and from prochlorococcus viruses. *Proceedings of the National Academy of Sciences of the United States of America*, *101*(30), 11013-11018.

Liu, Z., DeSantis, T. Z., Andersen, G. L., & Knight, R. (2008). Accurate taxonomy assignments from 16S rrna sequences produced by highly parallel pyrosequencers. *Nucleic Acids Research*, *36*(18), e120.

Livak, K. J. & Schmittgen, T. D. (2001). Analysis of relative gene expression data using real-time quantitative PCR and the 2-[delta][delta]CT method. *Methods*, *25*(4), 402-408.

Livny, J., Fogel, M. A., Davis, B. M., & Waldor, M. K. (2005). Srnapredict: An integrative computational approach to identify srnas in bacterial genomes. *Nucleic Acids Research*, *33*(13), 4096-4105.

Lo, I., Denef, V. J., VerBerkmoes, N. C., Shah, M. B., Goltsman, D., DiBartolo, G., et al. (2007). Strain-Resolved community proteomics reveals recombining genomes of acidophilic bacteria. *Nature*, *446*(7135), 537-541.

Longhurst, A. (1998). Ecological geography of the sea, 398 pp. *Acad. Press, San Diego, Calif.*

López-López, A., Bartual, S. G., Stal, L., Onyshchenko, O., & Rodríguez-Valera, F. (2005). Genetic analysis of housekeeping genes reveals a deep-sea ecotype of alteromonas macleodii in the mediterranean sea. *Environmental Microbiology*, *7*(5), 649-659.

Malmstrom, R. R., Coe, A., Kettler, G. C., Martiny, A. C., Frias-Lopez, J., Zinser, E. R., et al. (2010). Temporal dynamics of prochlorococcus ecotypes in the atlantic and pacific oceans. *The ISME Journal*, doi: 10.1038/ismej.2010.60.

Mandin, P. & Gottesman, S. (2010). Integrating anaerobic/aerobic sensing and the general stress response through the arcz small RNA. *The EMBO Journal*, *29*(18), 3094-3107.

Mardis, E. R. (2008). The impact of next-generation sequencing technology on genetics. *Trends in Genetics : TIG*, *24*(3), 133-141.

Margulies, M., Egholm, M., Altman, W. E., Attiya, S., Bader, J. S., Bemben, L. A., et al. (2005). Genome sequencing in microfabricated high-density picolitre reactors. *Nature*, *437*(7057), 376-380.

Marie, D., Partensky, F., Jacquet, S., & Vaulot, D. (1997). Enumeration and cell cycle analysis of natural populations of marine picoplankton by flow cytometry using the nucleic acid stain SYBR green I. *Applied and Environmental Microbiology*, *63*(1), 186-193.

Martens-Habbena, W., Berube, P. M., Urakawa, H., de la Torre, J. R., & Stahl, D. A. (2009). Ammonia oxidation kinetics determine niche separation of nitrifying archaea and bacteria. *Nature*, *461*(7266), 976-979.

Martinez, A., Bradley, A. S., Waldbauer, J. R., Summons, R. E., & DeLong, E. F. (2007).

Proteorhodopsin photosystem gene expression enables photophosphorylation in a heterologous host. *Proc. Natl Acad. Sci. USA*, *104*(13), 5590-5595.

Martinez, A., Tyson, G. W., & DeLong, E. F. (2010). Widespread known and novel phosphonate utilization pathways in marine bacteria revealed by functional screening and metagenomic analyses. *Environmental Microbiology*, *12*(1), 222-238.

Martiny, A. C., Coleman, M. L., & Chisholm, S. W. (2006). Phosphate acquisition genes in *prochlorococcus* ecotypes: Evidence for genome-wide adaptation. *Proc. Natl Acad. Sci. USA*, *103*(33), 12552-12557.

Martiny, A. C., Huang, Y., & Li, W. (2009). Occurrence of phosphate acquisition genes in prochlorococcus cells from different ocean regions. *Environmental Microbiology*, *11*(6), 1340-1347.

McAndrew, P. M., Bjorkman, K. M., Church, M. J., Morris, P. J., Jachowski, N., Williams, P. J. L., et al. (2007). Metabolic response of oligotrophic plankton communities to deep water nutrient enrichment. *Marine Ecology-Progress Series*, *332*, 63-75.

McCarren, J. & DeLong, E. F. (2007). Proteorhodopsin photosystem gene clusters exhibit co-evolutionary trends and shared ancestry among diverse marine microbial phyla. *Environmental Microbiology*, *9*(4), 846-858.

McCarren, J., Becker, J. W., Repeta, D. J., Shi, Y., Young, C. R., Malmstrom, R. R., et al. (2010). Microbial community transcriptomes reveal microbes and metabolic pathways associated with dissolved organic matter turnover in the sea. *Proceedings of the National Academy of Sciences of the United States of America*, *107*(38), 16420-16427.

McGrath, K. C., Mondav, R., Sintrajaya, R., Slattery, B., Schmidt, S., & Schenk, P. M. (2010). Development of an environmental functional gene microarray for soil microbial communities. *APPL ENVIRON MICROB*, *76*(21), 7161-7170.

Meyer, F., Paarmann, D., D'Souza, M., Olson, R., Glass, E. M., Kubal, M., et al. (2008). The metagenomics RAST server - a public resource for the automatic phylogenetic and functional analysis of metagenomes. *BMC Bioinformatics*, *9*(1), doi:10.1186/1471-2105-9-386.

Meyer, M. M., Ames, T. D., Smith, D. P., Weinberg, Z., Schwalbach, M. S., Giovannoni, S. J., et al. (2009). Identification of candidate structured rnas in the marine organism 'candidatus pelagibacter ubique'. *BMC Genomics*, *10*, 268.

Meyer, M. M., Roth, A., Chervin, S. M., Garcia, G. A., & Breaker, R. R. (2008). Confirmation of a second natural preq1 aptamer class in streptococcaceae bacteria. *RNA (New York, N.Y.)*, *14*(4), 685-695.

Mincer, T. J., Church, M. J., Taylor, L. T., Preston, C., Kar, D. M., & DeLong, E. F. (2007). Quantitative distribution of presumptive archaeal and bacterial nitrifiers in monterey bay and the north pacific subtropical gyre. *Environmental Microbiology*, *9*(5), 1162-1175.

Moll, P. R., Duschl, J., & Richter, K. (2004). Optimized RNA amplification using t7-rna-polymerase based in vitro transcription. *Analytical Biochemistry*, *334*(1), 164-174.

Moore, L. R. & Chisholm, S. W. (1999). Photophysiology of the marine cyanobacterium prochlorococcus: Ecotypic differences among cultured isolates. *Limnology and*

Oceanography, *44*(3), 628-638.

Moore, L. R., Ostrowski, M., Scanlan, D. J., Feren, K., & Sweetsir, T. (2005). Ecotypic variation in phosphorus-acquisition mechanisms within marine picocyanobacteria. *Aquatic Microbial Ecology*, *39*(3), 257-269.

Moore, L. R., Post, A. F., Rocap, G., & Chisholm, S. W. (2002). Utilization of different nitrogen sources by the marine cyanobacteria prochlorococcus and synechococcus. *Limnology and Oceanography*, *47*(4), 989-996.

Moreno-Paz, M. & Parro, V. (2006). Amplification of low quantity bacterial RNA for microarray studies: Time-Course analysis of leptospirillum ferrooxidans under nitrogen-fixing conditions. *Environmental Microbiology*, *8*(6), 1064-1073.

Morris, R. M., Vergin, K. L., Cho, J. C., Rappé, M. S., Carlson, C. A., & Giovannoni, S. J. (2005). Temporal and spatial response of bacterioplankton lineages to annual convective overturn at the bermuda atlantic time-series study site. *Limnology and Oceanography*, *50*(5), 1687-1696.

Noguchi, H., Park, J., & Takagi, T. (2006). Metagene: Prokaryotic gene finding from environmental genome shotgun sequences. *Nucleic Acids Research*, *34*(19), 5623-5630.

Orchard, E. D., Webb, E. A., & Dyhrman, S. T. (2009). Molecular analysis of the phosphorus starvation response in trichodesmium spp. *Environmental Microbiology*, *11*(9), 2400-2411.

Oz, A., Sabehi, G., Koblízek, M., Massana, R., & Béjà, O. (2005). Roseobacter-Like bacteria in red and mediterranean sea aerobic anoxygenic photosynthetic populations. *Applied and Environmental Microbiology*, *71*(1), 344-353.

Pace, N. R. (1997). A molecular view of microbial diversity and the biosphere. *Science*, *276*(5313), 734-740.

Padalon-Brauch, G., Hershberg, R., Elgrably-Weiss, M., Baruch, K., Rosenshine, I., Margalit, H., et al. (2008). Small rnas encoded within genetic islands of salmonella typhimurium show host-induced expression and role in virulence. *Nucleic Acids Research*, *36*(6), 1913-1927.

Parro, V., Moreno-Paz, M., & González-Toril, E. (2007). Analysis of environmental transcriptomes by DNA microarrays. *Environmental Microbiology*, *9*(2), 453-464.

Pham, V. D., Konstantinidis, K. T., Palden, T., & DeLong, E. F. (2008). Phylogenetic analyses of ribosomal dna-containing bacterioplankton genome fragments from a 4000 m vertical profile in the north pacific subtropical gyre. *Environmental Microbiology*, *10*(9), 2313-2330.

Poinar, H. N., Schwarz, C., Qi, J., Shapiro, B., MacPhee, R. D. E., Buigues, B., et al. (2006). Metagenomics to paleogenomics: Large-Scale sequencing of mammoth DNA. *Science*, *311*(5759), 392-394.

Poretsky, R. S., Bano, N., Buchan, A., LeCleir, G., Kleikemper, J., Pickering, M., et al. (2005). Analysis of microbial gene transcripts in environmental samples. *Applied and Environmental Microbiology*, *71*(7), 4121-4126.

Poretsky, R. S., Hewson, I., Sun, S., Allen, A. E., Zehr, J. P., & Moran, M. A. (2009). Comparative day/night metatranscriptomic analysis of microbial communities in the north pacific subtropical gyre. *Environmental Microbiology*, *11*(6), 1358-1375.

Pruesse, E., Quast, C., Knittel, K., Fuchs, B. M., Ludwig, W., Peplies, J., et al. (2007). SILVA: A comprehensive online resource for quality checked and aligned ribosomal RNA sequence data compatible with ARB. *Nucleic Acids Research*, *35*(21), 7188-7196.

Rachman, H., Lee, J. S., Angermann, J., Kowall, J., & Kaufmann, S. H. E. (2006). Reliable amplification method for bacterial RNA. *Journal of Biotechnology*, *126*(1), 61-68.

Ram, R. J., VerBerkmoes, N. C., Thelen, M. P., Tyson, G. W., Baker, B. J., Blake, R. C., et al. (2005). Community proteomics of a natural microbial biofilm. *Science*, *308*(5730), 1915-1920.

Reich, M., Liefeld, T., Gould, J., Lerner, J., Tamayo, P., & Mesirov, J. P. (2006). Genepattern 2.0. *Nature Genetics*, *38*(5), 500-501.

Ré, M. & Pavesi, G. (2007). Signal processing in comparative genomics. In F. Masulli, S. Mitra, & G. Pasi (Eds.), *Applications of fuzzy sets theory.* (pp. 544-50). New York: Springer.

Rich, V. I., Pham, V. D., Eppley, J., Shi, Y., & DeLong, E. F. (2010). Time-Series analyses of monterey bay coastal microbial picoplankton using a 'genome proxy' microarray. *Environmental Microbiology*, doi: 10.1111/j.1462-2920.2010.02314.x.

Riedel, T., Tomasch, J., Buchholz, I., Jacobs, J., Kollenberg, M., Gerdts, G., et al. (2010). Constitutive expression of the proteorhodopsin gene by a flavobacterium strain representative of the proteorhodopsin-producing microbial community in the north sea. *Applied and Environmental Microbiology*, *76*(10), 3187-3197.

Ripp, S. & Miller, R. V. (1997). The role of pseudolysogeny in bacteriophage-host interactions in a natural freshwater environment. *Microbiology (Reading, England)*, *143*(6), 2065-2070.

Ripp, S. & Miller, R. V. (1998). Dynamics of the pseudolysogenic response in slowly growing cells of pseudomonas aeruginosa. *Microbiology (Reading, England)*, *144 (Pt 8)*, 2225-2232.

Rippka, R., Coursin, T., Hess, W., Lichtlé, C., Scanlan, D. J., Palinska, K. A., et al. (2000). Prochlorococcus marinus chisholm et al. 1992 subsp. Pastoris subsp. Nov. Strain PCC 9511, the first axenic chlorophyll a2/b2-containing cyanobacterium (oxyphotobacteria). *Int J Syst Evol Microbiol*, *50 Pt 5*, 1833-1847.

Rocap, G., Distel, D. L., Waterbury, J. B., & Chisholm, S. W. (2002). Resolution of prochlorococcus and synechococcus ecotypes by using 16S-23S ribosomal DNA internal transcribed spacer sequences. *Applied and Environmental Microbiology*, *68*(3), 1180-1191.

Rocap, G., Larimer, F. W., Lamerdin, J., Malfatti, S., Chain, P., Ahlgren, N. A., et al. (2003). Genome divergence in two prochlorococcus ecotypes reflects oceanic niche differentiation. *Nature*, *424*(6952), 1042-1047.

Rodrigue, S., Malmstrom, R. R., Berlin, A. M., Birren, B. W., Henn, M. R., & Chisholm, S. W. (2009). Whole genome amplification and de novo assembly of single bacterial cells. *Plos*

ONE, *4*(9), e6864.

Rodriguez-Brito, B., Rohwer, F., & Edwards, R. A. (2006). An application of statistics to comparative metagenomics. *BMC Bioinformatics*, *7*, 162.

Rudd, K. E. (2000). Ecogene: A genome sequence database for escherichia coli K-12. *Nucleic Acids Research*, *28*(1), 60-64.

Rusch, D. B., Halpern, A. L., Sutton, G., Heidelberg, K. B., Williamson, S., Yooseph, S., et al. (2007). The sorcerer II global ocean sampling expedition: Northwest atlantic through eastern tropical pacific. *Plos Biology*, *5*(3), 398-431.

Sabehi, G., Loy, A., Jung, K. H., Partha, R., Spudich, J. L., Isaacson, T., et al. (2005). New insights into metabolic properties of marine bacteria encoding proteorhodopsins. *Plos Biology*, *3*(8), 1409-1417.

Said, N., Rieder, R., Hurwitz, R., Deckert, J., Urlaub, H., & Vogel, J. (2009). In vivo expression and purification of aptamer-tagged small RNA regulators. *Nucleic Acids Research*, *37*(20), e133.

Saito, M. A., Goepfert, T. J., & Ritt, J. T. (2008). Some thoughts on the concept of colimitation: Three definitions and the importance of bioavailability. *Limnology and Oceanography*, *53*(1), 276-290.

Schaber, J., Rispe, C., Wernegreen, J., Buness, A., Delmotte, F., Silva, F. J., et al. (2005). Gene expression levels influence amino acid usage and evolutionary rates in endosymbiotic bacteria. *Gene*, *352*, 109-117.

Schattner, P. (2002). Searching for RNA genes using base-composition statistics. *Nucleic Acids Research*, *30*(9), 2076-2082.

Schäfer, H., Servais, P., & Muyzer, G. (2000). Successional changes in the genetic diversity of a marine bacterial assemblage during confinement. *Archives of Microbiology*, *173*(2), 138-145.

Scherl, A., FranÁois, P., Bento, M., Deshusses, J. M., Charbonnier, Y., Converset, V., et al. (2005). Correlation of proteomic and transcriptomic profiles of staphylococcus aureus during the post-exponential phase of growth. *Journal of Microbiological Methods*, *60*(2), 247-257.

Schneider, J., Buness, A., Huber, W., Volz, J., Kioschis, P., Hafner, M., et al. (2004). Systematic analysis of T7 RNA polymerase based in vitro linear RNA amplification for use in microarray experiments. *BMC Genomics*, *5*(1), 29.

Selinger, D. W., Saxena, R. M., Cheung, K. J., Church, G. M., & Rosenow, C. (2003). Global RNA half-life analysis in escherichia coli reveals positional patterns of transcript degradation. *Genome Research*, *13*(2), 216-223.

Semon, M., Mouchiroud, D., & Duret, L. (2005). Relationship between gene expression and gc-content in mammals: Statistical significance and biological relevance. *Human Molecular Genetics*, *14*(3), 421-427.

Seymour, J. R., Ahmed, T., Durham, W. M., & Stocker, R. (2010). Chemotactic response of marine bacteria to the extracellular products of synechococcus and prochlorococcus.

Aquatic Microbial Ecology, *59*(2), 161-168.

Sharma, C. M., Hoffmann, S., Darfeuille, F., Reignier, J., Findeiss, S., Sittka, A., et al. (2010). The primary transcriptome of the major human pathogen helicobacter pylori. *Nature*, *464*(7286), 250-255.

Shi, Y., Tyson, G. W., & DeLong, E. F. (2009). Metatranscriptomics reveals unique microbial small rnas in the ocean's water column. *Nature*, *459*(7244), 266-269.

Silvaggi, J. M., Perkins, J. B., & Losick, R. (2006). Genes for small, noncoding rnas under sporulation control in *bacillus subtilis*. *Journal of Bacteriology*, *188*(2), 532-541.

Sogin, M. L., Morrison, H. G., Huber, J. A., Welch, D. M., Huse, S. M., Neal, P. R., et al. (2006). Microbial diversity in the deep sea and the underexplored "rare biosphere". *Proc. Natl Acad. Sci. USA*, *103*(32), 12115-12120.

Sowell, S. M., Wilhelm, L. J., Norbeck, A. D., Lipton, M. S., Nicora, C. D., Barofsky, D. F., et al. (2008). Transport functions dominate the SAR11 metaproteome at low-nutrient extremes in the sargasso sea. *The ISME Journal*, *3*(1), 93-105.

Sridhar, J. & Rafi, Z. A. (2007). Identification of novel genomic islands associated with small rnas. *In Silico Biology*, *7*(6), 601-611.

Steglich, C., Futschik, M. E., Lindell, D., Voss, B., Chisholm, S. W., & Hess, W. R. (2008). The challenge of regulation in a minimal photoautotroph: Non-Coding rnas in prochlorococcus. *Plos Genetics*, *4*(8), e1000173.

Steglich, C., Lindell, D., Futschik, M., Rector, T., Steen, R., & Chisholm, S. W. (2010). Short RNA half-lives in the slow-growing marine cyanobacterium prochlorococcus. *Genome Biology*, *11(5): R54.*

Steunou, A. S., Bhaya, D., Bateson, M. M., Melendrez, M. C., Ward, D. M., Brecht, E., et al. (2006). In situ analysis of nitrogen fixation and metabolic switching in unicellular thermophilic cyanobacteria inhabiting hot spring microbial mats. *Proc. Natl Acad. Sci. USA*, *103*(7), 2398-2403.

Steunou, A. S., Jensen, S. I., Brecht, E., Becraft, E. D., Bateson, M. M., Kilian, O., et al. (2008). Regulation of nif gene expression and the energetics of N-2 fixation over the diel cycle in a hot spring microbial mat. *Isme Journal*, *2*(4), 364-378.

Stewart, F. J., Ottesen, E. A., & DeLong, E. F. (2010). Development and quantitative analyses of a universal rrna-subtraction protocol for microbial metatranscriptomics. *The ISME Journal*, *4*(7), 896-907.

Storz, G. & Haas, D. (2007). A guide to small rnas in microorganisms editorial overview. *Current Opinion in Microbiology*, *10*, 93-95.

Storz, G., Altuvia, S., & Wassarman, K. M. (2005). An abundance of RNA regulators. *Annual Review of Biochemistry*, *74*, 199-217.

Sullivan, M. B., Lindell, D., Lee, J. A., Thompson, L. R., Bielawski, J. P., & Chisholm, S. W. (2006). Prevalence and evolution of core photosystem II genes in marine cyanobacterial viruses and their hosts. *Plos Biology*, *4*(8), e234.

Sullivan, M. B., Waterbury, J. B., & Chisholm, S. W. (2003). Cyanophages infecting the oceanic

cyanobacterium prochlorococcus. *Nature*, *424*(6952), 1047-1051.

Suttle, C. A. (1994). The significance of viruses to mortality in aquatic microbial communities. *Microbial Ecology*, *28*(2), 237-243.

Suttle, C. A. (2005). Viruses in the sea. *Nature*, *437*(7057), 356-361.

Suzuki, M. T., Preston, C. M., Béjà, O., de la Torre, J. R., Steward, G. F., & DeLong, E. F. (2004). Phylogenetic screening of ribosomal RNA gene-containing clones in bacterial artificial chromosome (BAC) libraries from different depths in monterey bay. *Microbial Ecology*, *48*(4), 473-488.

Tam, R. & Saier, M. H. (1993). Structural, functional, and evolutionary relationships among extracellular solute-binding receptors of bacteria. *Microbiological Reviews*, *57*(2), 320-346.

Tartar, A., Wheeler, M. M., Zhou, X., Coy, M. R., Boucias, D. G., & Scharf, M. E. (2009). Parallel metatranscriptome analyses of host and symbiont gene expression in the gut of the termite reticulitermes flavipes. *Biotechnology for Biofuels*, *2*, 25.

Tatusov, R. L., Galperin, M. Y., Natale, D. A., & Koonin, E. V. (2000). The COG database: A tool for genome-scale analysis of protein functions and evolution. *Nucleic Acids Research*, *28*(1), 33-36.

Temperton, B., Field, D., Oliver, A., Tiwari, B., Mühling, M., Joint, I., et al. (2009). Bias in assessments of marine microbial biodiversity in fosmid libraries as evaluated by pyrosequencing. *The ISME Journal*, *3*(7), 792-796.

Toledo-Arana, A., Dussurget, O., Nikitas, G., Sesto, N., Guet-Revillet, H., Balestrino, D., et al. (2009). The listeria transcriptional landscape from saprophytism to virulence. *Nature*, *459*(7249), 950-956.

Tolonen, A. C., Aach, J., Lindell, D., Johnson, Z. I., Rector, T., Steen, R., et al. (2006). Global gene expression of prochlorococcus ecotypes in response to changes in nitrogen availability. *Molecular Systems Biology*, *2*, 53.

Tringe, S. G. & Rubin, E. M. (2005). Metagenomics: DNA sequencing of environmental samples. *Nature Reviews Genetics*, *6*(11), 805-814.

Tripp, H. J., Schwalbach, M. S., Meyer, M. M., Kitner, J. B., Breaker, R. R., & Giovannoni, S. J. (2008). Unique glycine-activated riboswitch linked to glycine-serine auxotrophy in SAR11. *Environmental Microbiology*, *11*(1), 230-238.

Trotochaud, A. E. & Wassarman, K. M. (2005). A highly conserved 6S RNA structure is required for regulation of transcription. *Nature Structural & Molecular Biology*, *12*(4), 313-319.

Turnbaugh, P. J., Hamady, M., Yatsunenko, T., Cantarel, B. L., Duncan, A., Ley, R. E., et al. (2009). A core gut microbiome in obese and lean twins. *Nature*, *457*(7228), 480-484.

Turnbaugh, P. J., Ley, R. E., Mahowald, M. A., Magrini, V., Mardis, E. R., & Gordon, J. I. (2006). An obesity-associated gut microbiome with increased capacity for energy harvest. *Nature*, *444*(7122), 1027-1031.

Turnbaugh, P. J., Quince, C., Faith, J. J., McHardy, A. C., Yatsunenko, T., Niazi, F., et al.

(2010). Organismal, genetic, and transcriptional variation in the deeply sequenced gut microbiomes of identical twins. *Proceedings of the National Academy of Sciences of the United States of America*, *107*(16), 7503-7508.

Tyson, G. W., Chapman, J., Hugenholtz, P., Allen, E. E., Ram, R. J., Richardson, P. M., et al. (2004). Community structure and metabolism through reconstruction of microbial genomes from the environment. *Nature*, *428*(6978), 37-43.

Ueda, H. R., Hayashi, S., Matsuyama, S., Yomo, T., Hashimoto, S., Kay, S. A., et al. (2004). Universality and flexibility in gene expression from bacteria to human. *Proc. Natl Acad. Sci. USA*, *101*(11), 3765-3769.

Urich, T., Lanzén, A., Qi, J., Huson, D. H., Schleper, C., & Schuster, S. C. (2008). Simultaneous assessment of soil microbial community structure and function through analysis of the meta-transcriptome. *Plos ONE*, *3*(6), e2527.

Van Mooy, B. A. S. & Devol, A. H. (2008). Assessing nutrient limitation of prochlorococcus in the north pacific subtropical gyre by using an RNA capture method. *Limnology and Oceanography*, *53*(1), 78-88.

Vangelder, R. N., Vonzastrow, M. E., Yool, A., Dement, W. C., Barchas, J. D., & Eberwine, J. H. (1990). Amplified RNA synthesized from limited quantities of heterogeneous cdna. *Proc. Natl Acad. Sci. USA*, *87*(5), 1663-1667.

Venter, J. C., Remington, K., Heidelberg, J. F., Halpern, A. L., Rusch, D., Eisen, J. A., et al. (2004). Environmental genome shotgun sequencing of the sargasso sea. *Science*, *304*(5667), 66-74.

Vila-Costa, M., Rinta-Kanto, J. M., Sun, S., Sharma, S., Poretsky, R., & Moran, M. A. (2010). Transcriptomic analysis of a marine bacterial community enriched with dimethylsulfoniopropionate. *The ISME Journal*, 10.1038/ismej.2010.62.

Vogel, J. & Wagner, E. G. (2007). Target identification of small noncoding rnas in bacteria. *Current Opinion in Microbiology*, *10*(3), 262-270.

Vogel, J., Bartels, V., Tang, T. H., Churakov, G., Slagter-Jager, J. G., Huttenhofer, A., et al. (2003). Rnomics in *escherichia coli* detects new srna species and indicates parallel transcriptional output in bacteria. *Nucleic Acids Research*, *31*(22), 6435-6443.

von Wintzingerode, F., Göbel, U. B., & Stackebrandt, E. (1997). Determination of microbial diversity in environmental samples: Pitfalls of pcr-based rrna analysis. *FEMS Microbiology Reviews*, *21*(3), 213-229.

Walker, C. B., de la Torre, J. R., Klotz, M. G., Urakawa, H., Pinel, N., Arp, D. J., et al. (2010). Nitrosopumilus maritimus genome reveals unique mechanisms for nitrification and autotrophy in globally distributed marine crenarchaea. *Proc. Natl Acad. Sci. USA*, *107*(19), 8818-8823.

Wang, L., Feng, Z., Wang, X., Wang, X., & Zhang, X. (2010). Degseq: An R package for identifying differentially expressed genes from rna-seq data. *Bioinformatics*, *26*(1), 136-138.

Warnecke, F., Luginbühl, P., Ivanova, N., Ghassemian, M., Richardson, T. H., Stege, J. T., et al. (2007). Metagenomic and functional analysis of hindgut microbiota of a wood-feeding

higher termite. *Nature*, *450*(7169), 560-565.

Washietl, S., Hofacker, I. L., & Stadler, P. F. (2005). Fast and reliable prediction of noncoding rnas. *Proc. Natl Acad. Sci. USA*, *102*(7), 2454-2459.

Waters, L. S. & Storz, G. (2009). Regulatory rnas in bacteria. *Cell*, *136*(4), 615-628.

Weinbauer, M. G. (2004). Ecology of prokaryotic viruses. *FEMS Microbiology Reviews*, *28*(2), 127-181.

Weinberg, Z., Perreault, J., Meyer, M. M., & Breaker, R. R. (2009). Exceptional structured noncoding rnas revealed by bacterial metagenome analysis. *Nature*, *462*(7273), 656-659.

Wendisch, V. F., Zimmer, D. P., Khodursky, A., Peter, B., Cozzarelli, N., & Kustu, S. (2001). Isolation of escherichia coli mrna and comparison of expression using mrna and total RNA on DNA microarrays. *Analytical Biochemistry*, *290*(2), 205-213.

Whitman, W. B., Coleman, D. C., & Wiebe, W. J. (1998). Prokaryotes: The unseen majority. *Proc. Natl Acad. Sci. USA*, *95*(12), 6578-6583.

Williams, P. (1981). Microbial contribution to overall marine plankton metabolism: Direct measurements of respiration. *Oceanologica Acta. Paris*, *4*(3), 359-364.

Williamson, S. J., McLaughlin, M. R., & Paul, J. H. (2001). Interaction of the phihsic virus with its host: Lysogeny or pseudolysogeny? *Applied and Environmental Microbiology*, *67*(4), 1682-1688.

Wu, J. F., Sunda, W., Boyle, E. A., & Karl, D. M. (2000). Phosphate depletion in the western north atlantic ocean. *Science*, *289*(5480), 759-762.

Wu, J., Gao, W., Zhang, W., & Meldrum, D. R. (2010). Optimization of whole-transcriptome amplification from low cell density deep-sea microbial samples for metatranscriptomic analysis. *Journal of Microbiological Methods*, 10.1016/j.mimet.2010.10.018.

Yao, Z., Barrick, J., Weinberg, Z., Neph, S., Breaker, R., Tompa, M., et al. (2007). A computational pipeline for high- throughput discovery of cis-regulatory noncoding RNA in prokaryotes. *Plos Computational Biology*, *3*(7), e126.

Yergeau, E., Lawrence, J. R., Waiser, M. J., Korber, D. R., & Greer, C. W. (2010). Metatranscriptomic analysis of the response of river biofilms to pharmaceutical products, using anonymous DNA microarrays. *APPL ENVIRON MICROB*, *76*(16), 5432-5439.

Yooseph, S., Sutton, G., Rusch, D. B., Halpern, A. L., Williamson, S. J., Remington, K., et al. (2007). The sorcerer II global ocean sampling expedition: Expanding the universe of protein families. *Plos Biology*, *5*(3), 432-466.

Zemb, O., West, N., Bourrain, M., Godon, J. J., & Lebaron, P. (2010). Effect of a transient perturbation on marine bacterial communities with contrasting history. *Journal of Applied Microbiology*, *109*(3), 751-762.

Zhang, Y. & Gladyshev, V. N. (2008). Trends in selenium utilization in marine microbial world revealed through the analysis of the global ocean sampling (GOS) project. *Plos Genetics*, *4*(6), e1000095.

Zhou, J. & Thompson, D. K. (2002). Challenges in applying microarrays to environmental studies. *Current Opinion in Biotechnology*, *13*(3), 204-207.

Zinser, E. R., Coe, A., Johnson, Z. I., Martiny, A. C., Fuller, N. J., Scanlan, D. J., et al. (2006). Prochlorococcus ecotype abundances in the north atlantic ocean as revealed by an improved quantitative PCR method. *Applied and Environmental Microbiology*, *72*(1), 723-732.

Zinser, E. R., Lindell, D., Johnson, Z. I., Futschik, M. E., Steglich, C., Coleman, M. L., et al. (2009). Choreography of the transcriptome, photophysiology, and cell cycle of a minimal photoautotroph, *prochlorococcus. Plos ONE*, *4*(4), e5135.

Appendix A: Microbial community transcriptomes reveal microbes and metabolic pathways associated with dissolved organic matter turnover in the sea

Jay McCarren, Jamie W. Becker, Daniel J. Repeta, Yanmei Shi, Curtis R. Young, Rex R. Malmstrom, Sallie W. Chisholm, and Edward F. DeLong

McCarren, J., Becker, J.W., Repeta, D.J., Shi, Y., Young, C.R., Malmstrom, R.R., Chisholm, S.W., and DeLong, E.F. (2010). Microbial community transcriptomes reveal microbes and metabolic pathways associated with dissolved organic matter turnover in the sea. *Proc Natl Acad Sci* **USA** *107*, 16420-16427.

Microbial community transcriptomes reveal microbes and metabolic pathways associated with dissolved organic matter turnover in the sea

Jay McCarren[a,b], Jamie W. Becker[a,c], Daniel J. Repeta[c], Yanmei Shi[a], Curtis R. Young[a], Rex R. Malmstrom[a,d], Sallie W. Chisholm[a], and Edward F. DeLong[a,e,1]

Departments of [a]Civil and Environmental Engineering and [e]Biological Engineering, Massachusetts Institute of Technology, Cambridge, MA 02139; [c]Department of Marine Chemistry and Geochemistry, Woods Hole Oceanographic Institution, Woods Hole, MA 02543; [b]Synthetic Genomics, La Jolla, CA 92037; and [d]Joint Genome Institute, Walnut Creek, CA 94598

This contribution is part of the special series of Inaugural Articles by members of the National Academy of Sciences elected in 2008.

Contributed by Edward F. DeLong, August 2, 2010 (sent for review July 1, 2010)

Marine dissolved organic matter (DOM) contains as much carbon as the Earth's atmosphere, and represents a critical component of the global carbon cycle. To better define microbial processes and activities associated with marine DOM cycling, we analyzed genomic and transcriptional responses of microbial communities to high-molecular-weight DOM (HMWDOM) addition. The cell density in the unamended control remained constant, with very few transcript categories exhibiting significant differences over time. In contrast, the DOM-amended microcosm doubled in cell numbers over 27 h, and a variety of HMWDOM-stimulated transcripts from different taxa were observed at all time points measured relative to the control. Transcripts significantly enriched in the HMWDOM treatment included those associated with two-component sensor systems, phosphate and nitrogen assimilation, chemotaxis, and motility. Transcripts from *Idiomarina* and *Alteromonas* spp., the most highly represented taxa at the early time points, included those encoding TonB-associated transporters, nitrogen assimilation genes, fatty acid catabolism genes, and TCA cycle enzymes. At the final time point, *Methylophaga* rRNA and non-rRNA transcripts dominated the HMWDOM-amended microcosm, and included gene transcripts associated with both assimilatory and dissimilatory single-carbon compound utilization. The data indicated specific resource partitioning of DOM by different bacterial species, which results in a temporal succession of taxa, metabolic pathways, and chemical transformations associated with HMWDOM turnover. These findings suggest that coordinated, cooperative activities of a variety of bacterial "specialists" may be critical in the cycling of marine DOM, emphasizing the importance of microbial community dynamics in the global carbon cycle.

carbon cycle | marine | bacteria | metagenomics | metatranscriptomics

Microbial activities drive most of Earth's biogeochemical cycles. Many processes and players involved in these planetary cycles, however, remain largely uncharacterized, due to the inherent complexity of microbial community processes in the environment. Cycling of organic carbon in ocean surface waters is no exception. Though marine dissolved organic matter (DOM) is one of the largest reservoirs of organic carbon on the planet (1), microbial activities that regulate DOM turnover remain poorly resolved (2).

Marine DOM is an important substrate for heterotrophic bacterioplankton, which efficiently remineralize as much as 50% of total primary productivity through the microbial loop (3–6). Though some DOM is remineralized on short timescales of minutes to hours, a significant fraction escapes rapid removal. In marine surface waters, this semilabile DOM transiently accumulates to concentrations 2–3 times greater than are found in the deep sea (7), and represents a large inventory of dissolved carbon and nutrients that are potential substrates for marine microbes. Time-series analyses of semilabile DOM accumulation in temperate and subtropical upper ocean gyres show an annual cycle in DOC inventory with net accumulation following the onset of summertime stratification, and net removal following with deep winter mixing. In addition, multiyear time-series data suggest that surface-water DOM inventories have been increasing over the past 10–20 y (8). The ecological factors behind these seasonal and decadal DOC accumulations are largely unknown. Nutrient (N, P) amendments do not appear to result in a drawdown of DOC, and other factors such as the microbial community structure and the chemical composition of semilabile DOM have been invoked to explain the dynamics of the semilabile DOC reservoir (9, 10). Whatever the cause, the balance and timing of semilabile DOM remineralization are critical factors that influence the magnitude of DOM and carbon exported to the ocean's interior through vertical mixing.

There are significant challenges associated with characterizing and quantifying complex, microbially influenced processes such as DOM cycling in the sea. These challenges include inherent phylogenetic and population diversity and variability, the complexities of microbial community metabolic properties and interactions, and those associated with measuring microbial assemblage activities and responses on appropriate temporal and spatial scales. Past approaches have included measuring the bulk response of microbial communities to nutrient addition (e.g., community substrate incorporation or respiration), following changes in total or functional group cell numbers by microscopy or flow cytometry, or monitoring changes in relative taxa abundance, typically using rRNA-based phylogenetic markers. A number of field experiments (9–13) have indicated that specific shifts in microbial community composition might be linked to surface-water carbon utilization. However, the pure compound nutrient additions (such as glucose) frequently used in such field experiments (9, 11, 14, 15) may not well approximate the environmentally relevant chemical mixtures or compound concentrations present in naturally occurring DOM.

Though complications associated with direct experimentation on natural microbial communities limit our understanding of oceanic carbon cycling to some extent, significant insight into these processes have been recently reported. For example, Carlson et al. (10) showed differences among depth-stratified microbial communities that may be related to their ability to use semilabile DOM that

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

The authors declare no conflict of interest.

Freely available online through the PNAS open access option.

Data deposition: The sequences reported in this paper have been deposited in the GenBank database (accession nos. SRA020733.11 and HQ012268–HQ012278).

[1]To whom correspondence should be addressed. E-mail: delong@mit.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1010732107/-/DCSupplemental.

accumulates in ocean surface waters. In addition, phylogenetic analyses of time-series samples have identified some taxonomic groups that appear to be responsive to deep-water mixing events, which may be relevant to organic carbon cycling dynamics (16, 17).

To better define the processes and population dynamics associated with marine microbial DOM cycling in ocean surface waters, we performed controlled experiments using seawater microcosms amended with freshly prepared, naturally occurring DOM. High-molecular-weight DOM (HMWDOM, defined here as the size fraction >1,000 Da and <30,000 Da) was concentrated by ultrafiltration using a 1-nm membrane filter, followed by a second filtration step to remove viruses. Whole, unfiltered seawater was distributed into replicate microcosms (20 L each) that were incubated at in-situ temperatures and light intensities. The ambient concentration of dissolved organic carbon (DOC) in the unamended microcosms was 82 μM DOC, whereas the HMWDOM-amended microcosms contained 328 μM DOC, representing a 4-fold increase over ambient DOC concentration. Replicate control and experimental microcosms were sampled periodically over the course of a 27-h period.

The responses of microbial community members to HMWDOM addition over time were followed using flow cytometric, metagenomic, and metatranscriptomic analytical techniques. HMWDOM-induced shifts in microbial cell numbers, community composition, functional gene content, and gene expression were observed at each time point, as indicated by changes in the DOM-treated microcosms relative to an unamended control. The data indicated rapid and specific HMWDOM-induced shifts in transcription, metabolic pathway expression, and microbial growth that appear to be associated with HMWDOM turnover in ocean surface waters.

Results and Discussion

HMWDOM-Induced Cell Dynamics. Replicate microcosms were established immediately before sunrise and sampled over the course of 27 h to track the changes in microbial cell numbers, community composition, gene content, and gene expression in control vs. HMWDOM-treated microcosms. Though cell numbers in control microcosms remained constant over the time course of the experiment, the HMDOM-treated microcosm exhibited a ~50% increase in total cells within 19 h (Fig. 1*A*). Assuming a 50% growth efficiency, this HMWDOM-stimulated cell growth represents consumption of less than 1% of the total added DOC. Flow cytometry indicated that the majority (> 80%) of this increase in cells was attributable to the growth of a specific population of larger, high-DNA-content cells (Fig. 1*B*). The distinct flow cytometric signature of the HMWDOM-responsive population at the final time point allowed us to separate these large, high-DNA-content cells for further analyses (*SI Appendix*, Fig. S1). Large, high-DNA-content cells were isolated and collected via fluorescence-activated cell sorting and used to generate a SSU rRNA gene amplicon library. Near full-length rRNA gene sequences from the sorted cells recovered were all affiliated with the phylum Proteobacteria, falling into one of three clades (Fig. 1*C*). One subset of the flow-sorted cell population contained Alphaproteobacteria, closely related to *Thalassobius* isolates within the family Rhodobacteraceae. The remaining rRNA genes from the cell-sorted population were derived from Gammaproteobacteria, with one subset most closely related to *Alteromonas* isolates, and a second subset most similar to *Methylophaga* isolates within the order Thiotrichales.

Taxon-Specific Patterns of rRNA Gene and rRNA Representation in Control vs. HMWDOM-Treated Metagenomic and Metatranscriptomic Datasets. Community genomic DNA samples from T_0 and T_{27hrs} were pyrosequenced on the Roche 454 FLX platform, yielding ≈500,000 reads per sample (Table 1). Though SSU rDNA genes represent a small fraction (~1%) of the total genomic pyrosequencing reads, sufficient data (~500–750 individual reads) was available for phylogenetic analyses, which avoids PCR bias, and other artifacts associated with PCR amplicon "pyrotag" libraries (18–20). Classification of these of rRNA genes (*Methods*) provided

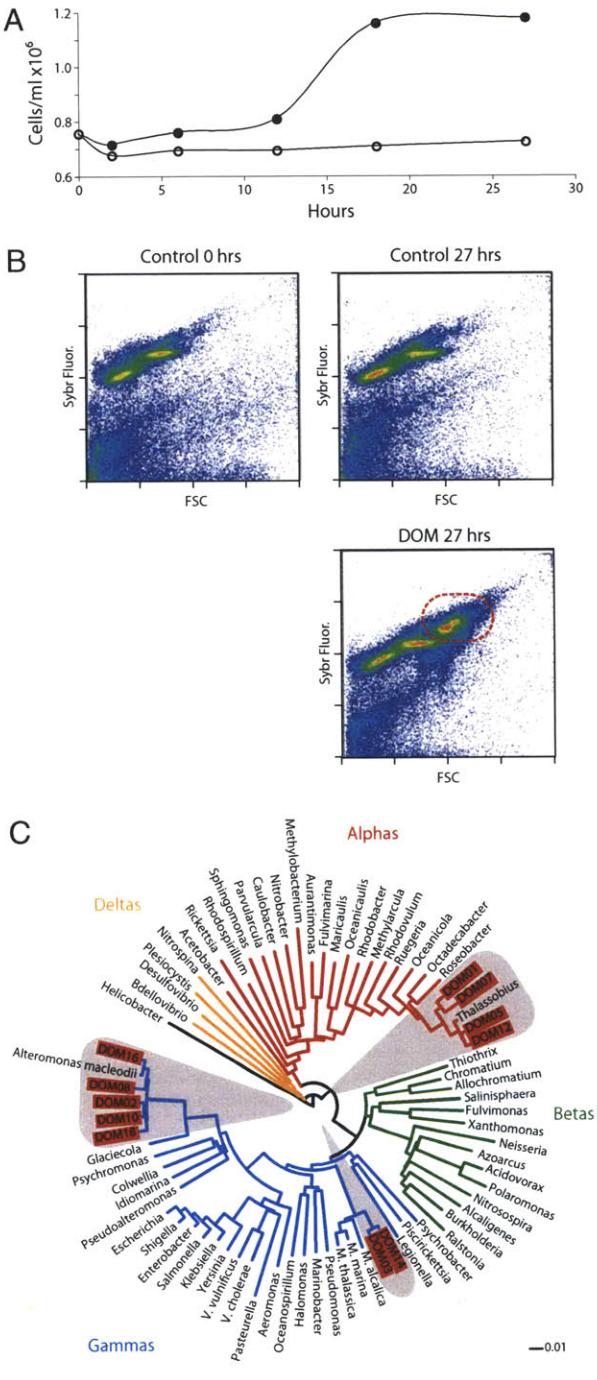


Fig. 1. Dynamics of microbial populations during 27-h microcosm incubations. (*A*) Flow cytometric counts of microbial cells from control (○) and DOM-amended (●) treatments. Samples displayed in *B* highlighted in red. (*B*) Flow cytometry scatterplots from selected samples show little change in the distribution of cell size [as measured by forward scatter (FSC)] and DNA content (SYBR fluorescence) of control samples from beginning to end of the experiment, whereas most of the increase in cell numbers observed in the DOM-amended treatment can be attributed to the appearance of larger, high-DNA-content cells (circled in red). (*C*) Weighted neighbor-joining tree of selected SSU SSU rDNA sequences from proteobacterial type strains and the sequences obtained from flow cytometric sorting of the larger, higher-DNA-content population of cells present after DOM amendment. The sequences obtained from the flow-sorted population are restricted to three specific taxonomic clades: Rhodobacteraceae, Methylophaga, and Alteromonas.

Table 1. Number of pyrosequences analyzed in control and treatment DNA and cDNA libraries

Treatment	Sample	0 h	2 h	12 h	27 h
Control	DNA	557,099	NA	NA	422,666
	cDNA	505,075	221,751	470,578*	514,670
	(non rRNA)	(18,345)	(12,658)	(12,934)	(18,078)
+DOM	DNA	NA	NA	NA	526,681
	cDNA	NA	230,376	251,690	751,284
	(non rRNA)	NA	(14,762)	(15,748)	(42,689)

*One of two technical replicate sequencing runs for this sample contained a spuriously high representation of a single sequence (~4.2% of reads) not present in the other replicate sequencing run. These nearly perfect duplicate reads (>99% nucleotide identity and read-length difference of <5 bp) were removed before subsequent analysis.

an overview of microbial community composition over the course of the experiment (Fig. 2*A*, inner rings). As expected, typically abundant planktonic bacterial taxa such as *Pelagibacter* (Rickettsiales) and *Prochlorococcus* (Cyanobacteria) were highly represented (Fig. 2*A* and *SI Appendix*, Fig. S2). The community composition of the control microcosm did not change substantially from the beginning to the end of the experiment. In contrast, the representation of several taxonomic groups increased in the HMWDOM-amended microcosm over the 27-h incubation. Three specific gammaproteobacterial groups—the families Idiomar-

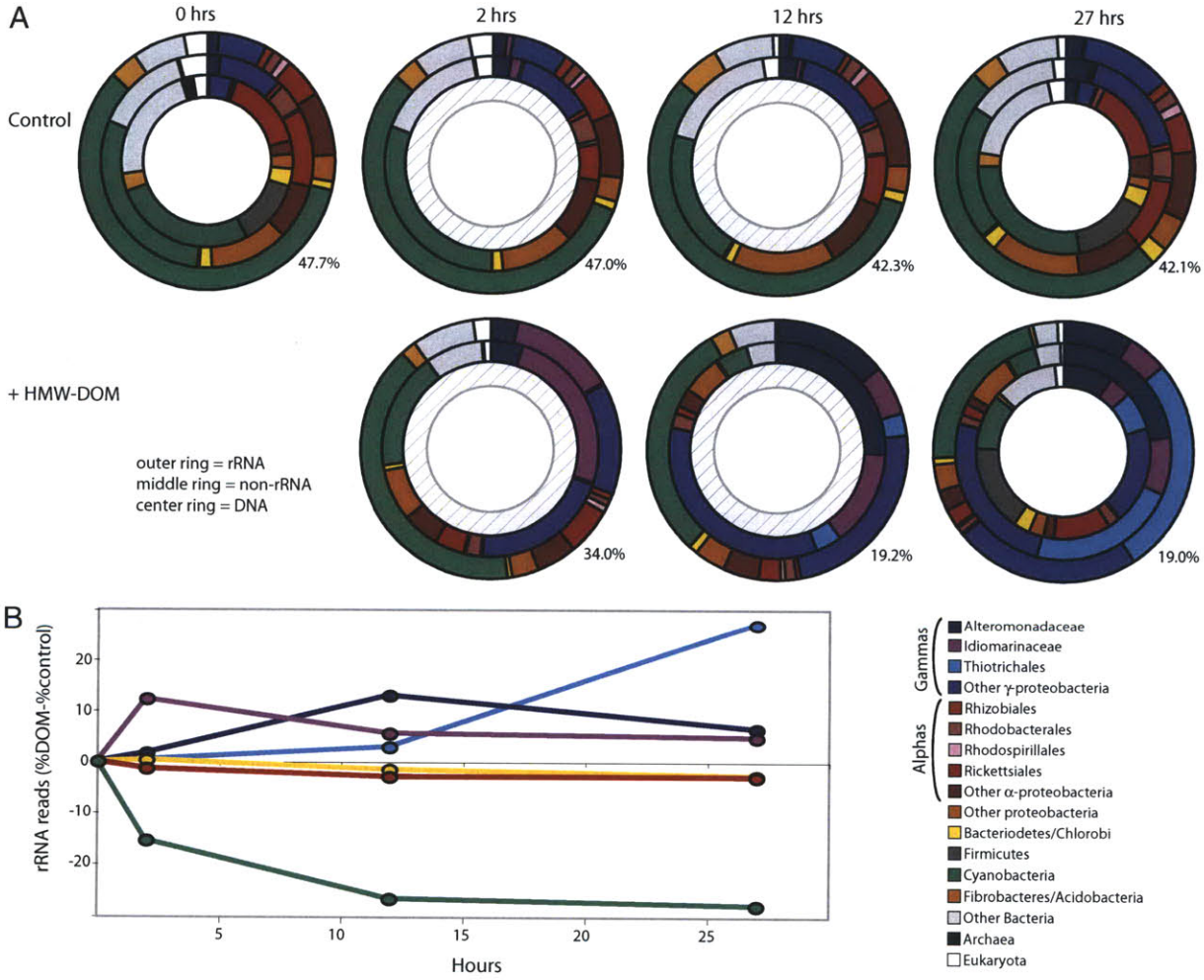


Fig. 2. Microbial community composition assessed by taxonomic classification of metagenomic and metatranscriptomic sequence reads. (*A*) SSU rRNA reads (outer ring) and non-rRNA reads (middle ring) from metatranscriptomic datasets as well as those reads from metagenomic datasets identified as SSU rDNA reads (center ring). Only taxonomic groups that represent >1% of total reads in at least one dataset have been included with all other groups binned together with unassigned reads. In some instances, reads can only be confidently assigned to broad class- and order-level taxonomic groups and are labeled as such. For mRNA datasets, some reads have no significant blast hits, the percentage of which is noted beside each sample. (*B*) Tracking the changes in community composition by comparing the difference between the DOM-amended treatment and control reveals distinct taxonomic groups responding at each time point. Only taxonomic groups showing more than ±2% change are plotted.

inaceae and Alteromonadaceae (both of which fall in the order Alteromonadales) and the order Thiotrichales—all increased in rRNA gene representation following HMWDOM amendment (Fig. 2 *A* and *B* and *SI Appendix*, Fig. S2). Two of these HMWDOM-stimulated groups (Alteromonadaceae and Thiotrichales) corresponded to the same dominant groups found in the FACS-sorted, high-DNA-containing cell populations (Fig. 1). The Rhodobacteraceae group that was recovered in the flow-sorted population did not, however, show a corresponding rRNA enrichment in the HMWDOM-treated metagenomic or metatranscriptomic datasets. These alphaproteobacteria may simply represent a background population of cells that were sorted along with the DOM-stimulated gammaproteobacteria because their flow cytometric signal overlapped with the large, high-DNA-content cell fraction.

Analyses of metagenomic sequence reads yields information on the relative representation of taxonomic groups, but not absolute cell numbers. Though cyanobacteria represented more than a quarter of all SSU rRNA genes throughout the time course of the experiment in the control microcosm, in the HMWDOM treatment they comprised only 10% of the rRNA sequence reads by 27 h. Enumeration of *Prochlorococcus* cells via flow cytometry indicated, however, that absolute *Prochlorococcus* cell numbers changed by less than 1% in the HMWDOM-amended microcosm. The changes in community composition observed in the metagenomic datasets therefore appear due to the growth of specific population members (in particular, Alteromonadaceae and Thiotrichales) and not to the disappearance of other dominant groups.

Compared with SSU rDNA reads from metagenomic DNA datasets, pyrosequencing of total community cDNA yielded orders of magnitude more total rRNA sequences that could be similarly classified taxonomically (Fig. 2*A*, outer rings). [The cDNAs in this study were not subjected to upstream rRNA subtraction procedures that have been reported in other metatranscriptomic studies (21–23).] In contrast to rRNA gene abundance in the DNA, rRNA in the cDNA pool reflects the cellular abundance of specific phylogentic groups, as well as their cellular rRNA copy numbers. For example, the rRNAs of several groups (e.g., Rickettsiales, Firmicutes, and Archaea) were less abundant in the cDNA datasets in comparison with their corresponding genes in the genomic DNA dataset (Fig. 2 and *SI Appendix*, Fig. S2). Conversely, cyanobacterial rRNAs were more highly represented in the cDNA than the corresponding rRNA genes in the DNA (Fig. 2 and *SI Appendix*, Fig. S2). Similarly, in the 27 h post-HMWDOM amendment, the Thiotrichales comprised nearly one-third of all SSU rRNA sequences in the cDNA, but represented less than 8% of all SSU rRNA genes in the DNA of the same sample.

Taxon-Specific Responses to HMWDOM Addition Inferred from Functional Gene Transcript Abundance. Taxonomic classification of non-rRNA transcripts from cDNA datasets (Fig. 2*A*, middle ring; *Methods*) generally paralleled the trends observed for rRNA taxon abundance, indicating parallel responses in both functional gene transcript and rRNAs (Fig. 2). Two exceptions to this correspondence were observed: cyanobacterial rRNA sequences were present in much greater abundance than non-rRNA cyanobacterial transcripts at all time points in both the control and the HMWDOM treatment. Conversely, Idiomarinaceae and Alteromondaceae were underrepresented in rRNAs, relative to non-rRNA transcripts present in the HMWDOM-treated microcosm cDNAs.

Distinct shifts in the cDNAs of specific subpopulations occurred in response to HMWDOM addition. Though the control remained virtually unchanged throughout the experiment, at each time point following HMWDOM addition, a different taxonomic group dominated the cDNA pool for both rRNA and non-rRNA transcripts (Fig. 2 *A* and *B*). Two hours post-HMWDOM amendment, Idiomarinaceae sequences represented nearly 13% of all rRNA sequences in the cDNAs from the HMWDOM treatment, though they remained less than 1% of the total rRNA sequences in all control cDNAs. By 12 h, the abundance of Idiomarinaceae rRNA sequences in the HMWDOM treatment receded closer to control values, whereas Alteromonadaceae rRNA sequences in the transcript pool rose to 15% of the total rRNAs relative to the control (Fig. 2*B*). Similarly, by the end of the experiment, Alteromonadaceae rRNA sequences decreased in relative abundance compared with earlier time points, when Thiotrichales-like rRNA represented the most abundant rRNAs. Strikingly, though Thiotrichales-like rRNAs represented approximately one-third of the total rRNA sequences in cDNA at the final HMWDOM-treated time point, Thiotrichales never represented more than 0.04% of in any of the controls at all time points.

Idiomarinaceae and Alteromonadaceae are closely related families within the order Alteromonadales (24). Because these closely related taxa were differentially represented at two different time points in the HMWDOM treatment, we searched for potential differences in their functional gene transcript representation at different times. All sequence reads having a best match to the full genome sequence of these two dominant taxa [*Idiomarina loihiensis* (25) and *Alteromonas macleodii* (26)] were analyzed separately for each taxonomic bin (*SI Appendix*, Tables S1 and S2). There were many similarities in the distribution of cDNA reads of functional gene categories between the two taxa. Examination of the 2-h and 12-h HMWDOM microcosm time points for Idiomarinaceae and Alteromonadaceae, respectively, indicated that transcript representation for many nutrient acquisition genes were similarly abundant within both taxonomic groups at the two different time points. An outer membrane receptor for a TonB-associated iron transporter was among the most abundant transcripts for both Idiomarinaceae and Alteromonadaceae. Similarly, the three genes require for the glutamine synthase cycle involved in nitrogen assimilation were abundant in each taxonomic bin. Genes involved in fatty acid catabolism were abundant in both Idiomarinaceae and Altermonadaceae bins (*SI Appendix*, Tables S1 and S2). Additionally, the two enzymes specific for the glyoxylate cycle (isocitrate lyase and malate synthase), which could use acetyl-CoA output by the β-oxidation of fatty acids, were abundant in both bins. One striking difference between the two different Alteromonadales cDNA bins was the high representation of one gene, triacylglycerol lipase (10-fold more abundant in treatment than control), found only among Idiomarinaceae-like reads. Interestingly, triacylglycerol lipase reads were virtually absent from reads assignable to the Alteromonadaceae bin.

The taxonomic groups that appeared most responsive to HMWDOM addition comprised only a small fraction of the starting microbial community. In contrast, transcripts from typically more dominant taxa such as *Pelagibacter* and *Prochlorococcus* decreased in relative abundance in the HMWDOM treatment over time. Additionally, because the differences in transcript abundance between control and treatment were small for *Prochlorococcus* and *Pelagibacter*, our sequencing depth allowed the detection of only a few significantly different transcripts between controls and treatments (*SI Appendix*, Figs. S3 and S4). Only seven *Pelagibacter* ORFs were identified as having statistically significant changes in transcript abundance ($P < 0.001$; *Methods*) in the HMWDOM-treated sample vs. the control (*SI Appendix*, Fig. S3). This small number of transcriptionally responsive ORFs (within our detection limits) was consistent with the hypothesis that *Pelagibacter* has a relatively small genome and streamlined regulatory network (27) and so may be less responsive to large fluctuations in ambient nutrient concentrations. The absolute *Pelagibacter* cell numbers appear to have increased slightly over the course of incubation in the treatment relative to the control, as evidenced by its higher gene abundances in the treatment relative to *Prochlorococcus* (whose absolute cell numbers remained constant as determined by flow cytometry; Fig. 2). The enrichment of transcripts encoding DNA-directed RNA polymerase and methionine biosynthesis protein (*SI Appendix*, Fig. S3) may indicate some utilization of some fraction of HMWDOM by *Pelagibacter*

cells to obtain reduced sulfur for the biosynthesis of sulfur-containing amino acids (28). The depletion of proteorhodopsin transcripts in the treatment at the final time point (*SI Appendix*, Fig. S3) suggested a potentially diminished requirement for proteorhodopsin phototrophy, with the increase in carbon availability. For *Prochlorococcus*, most of the significantly different transcripts were depleted in the treatment relative to the control at the earlier time points, whereas a few transcripts were enriched at the final time point. Several of these treatment-stimulated *Prochlorococcus* transcripts appeared to be involved with cellular repair processes, including oxidative damage protection and protein folding (*SI Appendix*, Fig. S4).

Small RNAs. Thirty putative sRNA (psRNA) clusters comprising >100 reads were identified, 20 of which showed statistically significant differences in abundance between the treatment and control for one or more time points (*SI Appendix*, Fig. S5). Based on the Rfam 10.0 database (http://rfam.sanger.ac.uk/), five clusters were identified as transfer-messenger RNA (tmRNA), and one was RNaseP RNA. Notably, all but one tmRNA cluster was overrepresented in the treatment, in part reflecting increases in specific taxa in the treatment vs. control (Fig. 1). For instance, cluster 7 tmRNA, which was overrepresented at 2 h, was most closely related to Idiomarinaceae, whereas *Methylophaga*-like cluster 9 tmRNA was enriched at later time points. Several psRNA clusters mapped into previously reported abundant psRNA groups found in microbial community transcripts sampled from the water column at Station ALOHA (29) (*SI Appendix*, Fig. S5). Five apparently different psRNA clusters (cluster 2, 3, 4, 8, and 14) were adjacent to genes encoding class II fumarate hydratase, an enzyme that catalyzes the reversible hydration/dehydration of fumarate to S-malate in the tricarboxylic acid cycle. To test the possibility that these clusters belonged to the same group but did not merge due to stringent clustering method, we performed pairwise alignment analysis among representative sequences of these five clusters (*SI Appendix*, Fig. S6). Only cluster 3 and cluster 14 merged (based on high sequence identity in the alignment at the end of both sequences), confirming that several divergent psRNA species, all adjacent to fumarate hydratase genes, were enriched in response to HMWDOM addition.

Global trends in functional gene transcript abundances in the HMWDOM treatment vs. control. All non-rRNA cDNA sequences were compared with NCBI-nr, KEGG (30), and GOS protein clusters databases (31) using BLASTX (32). We focused in particular on quantifying KEGG ortholog abundances in the HWM DOM-treated microcosm relative to the unamended controls across all time points (*SI Appendix*, Tables S3–S6).

Among all of the controls (0 h, 2 h, 12 h, and 27 h), only a few orthologs exhibited significant changes between time points ($n = 43$; *SI Appendix*, Table S3). Among these significantly different orthologs, about half were due to differences between the initial time point (0 h) and the other controls. In contrast, a larger number of orthologs exhibited differences in abundance between the pooled controls and the HMWDOM treatment (*SI Appendix*, Tables S4–S6). At 2 h post-HMWDOM addition, 67 KEGG orthologs exhibited differences from the control, with 58 of those enriched in the treatment vs. pooled controls (detectable effect sizes of enriched orthologs: 2.0- to 550-fold change; *SI Appendix*, Table S4). At 12 h, 221 differences were apparent, and 200 of those were enriched in the treatment vs. controls (detectable effect sizes of enriched orthologs: 2.3- to 2,200-fold change; *SI Appendix*, Table S5). At 27 h, 390 differences were detected, and 311 of those orthologs were enriched in the treatment (detectable effect sizes of enriched orthologs: 1.6- to 1,100-fold change; *SI Appendix*, Table S6).

Significantly enriched transcripts in the HMWDOM treatment included those encoding enzymes in KEGG pathways for carbohydrate, nitrogen, methane, sulfur, and fatty acid metabolic genes. Numerous transcripts associated with signal transduction and membrane transport pathways were also enriched in the HMWDOM treatment. Amino acid and nucleotide metabolism were also enriched in the HMWDOM addition microcosms, as were transcripts encoding enzymes involved in transcription and translation. The effect for all of these categories, however, was much more pronounced for the 12- and 27-h post-HMWDOM treatments than for the 2-h treatment. This is apparently due to the fact that the predominant DOM-responsive taxa were initially low in numbers, but increased in both cell density and transcriptional activity over the time course of the experiment.

At 12 h in the HMWDOM microcosm a variety of two-component sensor systems and several transporters were overrepresented. Particularly abundant were genes involved in nutrient acquisition. Specifically, both the components of the phosphate two-component sensor system (phoB, phoR, phoA, and OmpR phoB) as well as all components of the ABC transporter for phosphate (pstS, pstC, pstA, and pstB) were overrepresented at 12 and 27 h post-HMWDOM addition. At 27 h post-HMWDOM addition, members of several two-component sensor systems are enriched, including those associated with glucose (BarA, UvrY, CsrA), glucose-6-P (UhpB), nitrogen (GlnL, GlnG), C4-dicarboxylate (YfhK, YfhA), redox state of the quinone pool (ArcA), misfolded proteins (CpxR), carbon storage (BarA, UvrY, CsrA), and bacterial flagellar chemotaxis (CheA, CheV, CheY). Flagellar biosynthesis-associated transcripts were also similarly enriched, with 18 of 42 KOs associated with flagellar biosynthesis more the 4-fold more abundant in the amended microcosm relative to controls.

Transcripts encoding components of the GS/GOGAT pathway (glutamine and glutamate synthesis) were also significantly enriched in the HMWDOM treatment. Nitrogen two-component systems enriched in the DOM treatment transcript pool (GlnL, GlnG) typically sense nitrogen limitation via the intracellular glutamine pool and respond to nitrogen limitation by activating glutamate metabolism (33), which is consistent with the observed elevated GS/GOGAT transcript levels. Other enzymes in the nitrogen pathway, however, appeared relatively unchanged except for aminomethyltransferase (involved in glycine synthesis), which was less prevalent in the HMWDOM treatment. [Transcripts for one specific family of Amt family ammonium transporters from *Prochlorococcus* were significantly depleted in the HMWDOM treatment (*SI Appendix*, Fig. S4)]. Similar to the signatures of nitrogen limitation, the prevalence of the OmpR family phosphate two-component system, and the enrichment of a PIT family inorganic phosphate transporter, suggested that over the course of the experiment, the HMWDOM microcosm community was experiencing nitrogen and phosphate limitation as a consequence of the elevated DOC levels relative to the control.

Transcripts associated with sulfur-metabolizing enzymes were enriched in the HMWDOM treatment at the final time point and included enzymes associated with sulfate metabolism, and serine metabolism. Serine metabolism produces acetate that potentially could be shunted into the reductive carboxylate cycle, also enriched in the DOM treatment. Transcripts encoding three enzymes of the fatty acid metabolism pathway were also enriched in the HMWDOM treatment, as well as those encoding a short-chain fatty acid transporter. Furthermore, fatty acid biosynthesis pathway transcripts were significantly depleted in the HMWDOM treatment, suggesting a potential shift to catabolic metabolism of fatty acid-like molecules in the HMWDOM treatment. At the first time point, just 2 h postamendment, the two most enriched transcripts that corresponded to KEGG orthologs were triacylglycerol lipase and acyl-CoA dehydrogenase (50-fold and 109-fold, respectively). These enzymes catalyze two early steps in the catabolism of triacylglycerols (TAGs). These signals may be the result of cell wall material copartitioning in the HMWDOM concentrate, or the tendency of lipid compounds to associate with HMWDOM concentrates (34).

Methylophaga species were the most highly represented single taxon in both rRNA and functional gene transcripts in the

HMWDOM microcosm at the final time point. Consistent with this observation, two key enzymes involved in the ribulose monophosphate (RuMP) pathway, hexulose-6-phosphate synthase and 6-phospho-3-hexuloisomerase, were also highly abundant in the amended microcosm (eighth and second most abundant, respectively) while remaining undetected in the control. The cyclical RuMP pathway is an assimilatory pathway that is widespread in bacteria, functioning as a pathway for formaldehyde fixation and detoxification. In the first two reactions in this pathway, formaldehyde is condensed with ribulose-5 phosphate, which is then isomerized to fructose-6-phosphate. Moreover, gene transcripts for the enzymes encoding many of the steps in this pathway were enriched by the end of this experiment (Fig. 3) and increased over the time course of the experiment (Poisson ANOVA; *SI Appendix*, Table S7). Though a large variety of one-carbon compounds are processed through the RuMP pathway, all methyltrophic pathways share formaldehyde as a common entry point. Formaldehyde can also be oxidized to CO_2 via several routes, and several of the enzymes involved in these dissimilatory pathways were also abundant in the amended treatment (Fig. 3), particularly those associated with the tetrahydromethanopterin-dependent pathway. In total, the data reflected the enrichment of pathways for both assimilatory and dissimilatory single-carbon compound utilization, which coincided with the appearance of an actively growing *Methylophaga* population in the HMWDOM treatment (Figs. 2 and 3).

Conclusions

Semilabile DOM may support up to 40% of marine bacterial carbon demand (35, 36), yet little is known about the specific microorganisms and metabolic pathways responsible for its degradation and transformation in the ocean's water column. There is growing evidence that microbial transformation of semilabile DOM renders DOM less and less labile, further increasing accumulation in oligotrophic gyres and ultimately leading to export as refractory DOM (36). Microbial population dynamics and metabolic processes are therefore central to understanding the cycling of DOM in the sea.

In this study, short-term incubation of bacterial populations from surface seawater with naturally occurring HMWDOM from the same environment revealed specific shifts in microbial cells, rRNAs, and DOM-responsive gene transcripts relative to unamended controls. Cell numbers nearly doubled specifically in response to HMWDOM. Flow sorting and rRNA gene and transcript abundances consistently indicated the stimulation of several phylogenetic groups within the Alteromondales (*Idiomarina* and *Alteromonas* sp.) and Thiotrichales (*Methylophaga* sp.). Analysis of microbial cDNA abundances over time via pyrosequencing revealed that 2 h after DOM addition, close relatives of *Idioma-*

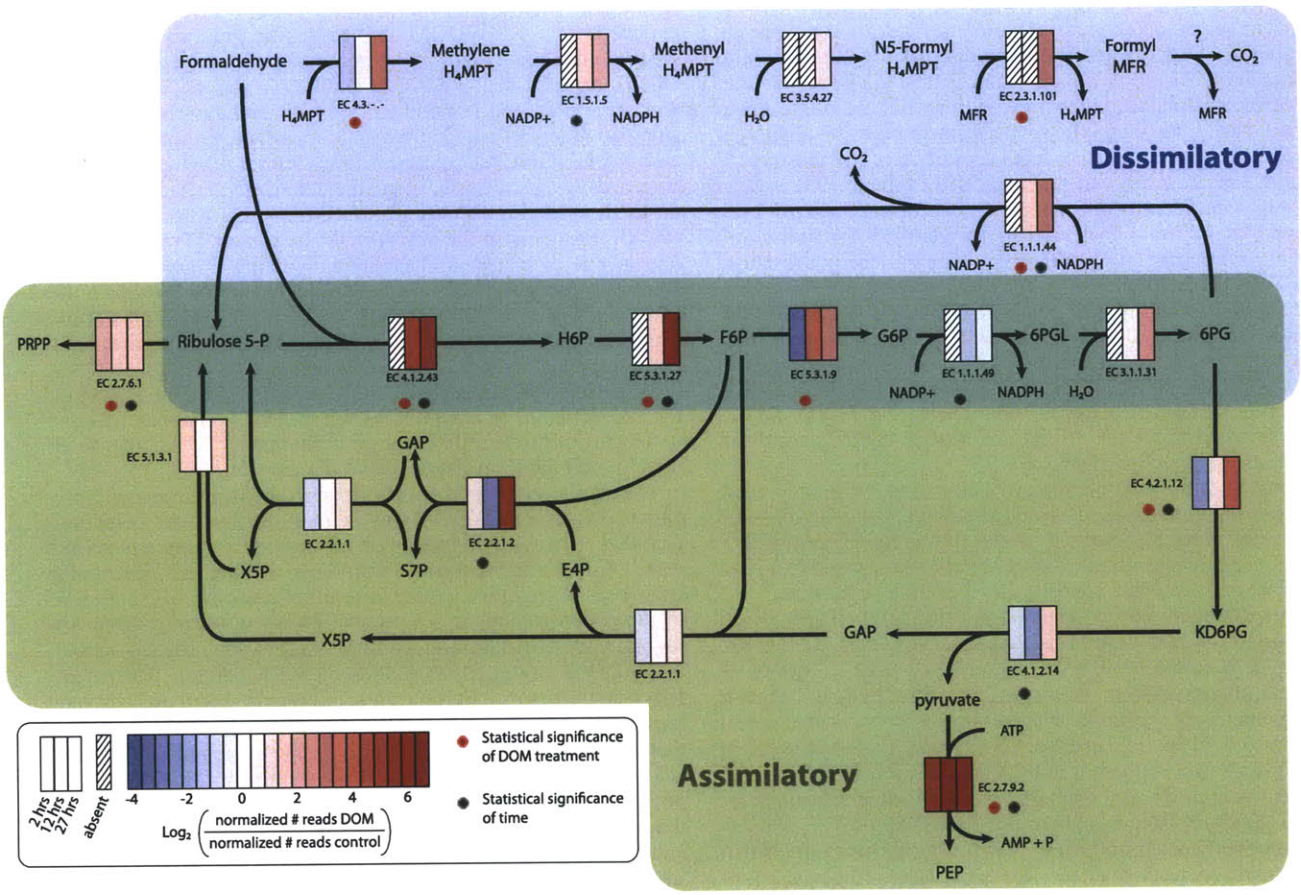


Fig. 3. Diagram of representative dissimilatory and assimilatory methylotrophic pathways and enzymes that show increased transcript abundance following DOM amendment. A KEGG ortholog-based expression ratio comparing normalized abundances of reads present in the DOM-amended treatment with those from an untreated control at 2, 12, and 27 h following DOM addition. Asterisks mark those enzymes showing statistically significantly differences in transcript abundance relative to time and/or unamended control (*SI Appendix*, Table S7). H_4MPT, tetrahydromethanopterin; MFR, formylmethanofuran; H6P, hexulose-6-phosphate; F6P, fructose-6-phosphate; 6PGL, 6-phosphogluconolactone; 6PG, 6-phosphogluconate; KD, ketodeoxy; PEP, phosphoenolpyruvate; GAP, glyceraldehyde phosphate; E4P, erythrose-4-phosphate; X5P, xylulose-5-phosphate; S7P, sedoheptulose-5-phosphate; PRPP, phosphoribosyl diphosphate.

rina sp. were stimulated by HMWDOM. In apparent microbial succession, a few hours later, *Alteromonas macleodiii*-like rRNAs and mRNAs increased dramatically relative to the unamended control. After 27 h, the same indicators showed that *Methylophaga* sp. (order Thiotrichales) predominated. We interpret this succession as a specific metabolic sequence and successional cascade that reflects sequential processing and degradation of specific components within HMWDOM. Analyses also indicated that 27 h post-DOM addition, both the dissimilatory and assimilatory single-carbon compound utilization pathways were highly expressed, coincident with the appearance and high abundance of *Methylophaga* sp. at the final time point.

The data indicate several specific groups of bacteria that appear to operate in succession and synergy to catalyze the turnover of naturally occurring HMWDOM in the marine environment. These findings may reflect regular (and predictable) metabolic cascades and community succession patterns that in part regulate the transformation and turnover of naturally occurring semilabile DOM. Furthermore, our findings are suggestive of some of the chemical attributes and degradation patterns of naturally occurring DOM. In previous chemical analyses, about 15% of DOM carbohydrate has been shown to consist of methyl sugars (37, 38). Our present findings suggest that Alteromonadales (specifically, *Idiomarina* spp. and *Alteromonas macleodii*) might be metabolizing semilabile DOM methyl sugars to methanol or formaldehyde, and carbon dioxide, among other products. The methanol and/or formaldehyde produced could be further oxidized and incorporated by *Methylophaga* sp. in the terminal portion of this aerobic food chain. Such a specific carbon compound-driven syntrophy has rarely been observed in aerobic microbial consortia. Although confirmation awaits further experimentation and chemical analyses, if correct, DOM methyl sugar metabolism might provide a partial explanation for the ubiquitous presence of methylotrophs in open-ocean and coastal environments (12, 39–42).

In summary, the experimental metatranscriptomic approach described here is beginning to reveal metabolic pathways and microbial taxa involved in the chemical transformation and turnover of naturally occurring marine DOM. These techniques can be used to track a variety of microbial processes in the environment, and set the stage for future inquiries on the nature and details of microbial community environmental responses and dynamics in situ. In this study, we gained detailed perspective on microbial community dynamics and metabolism associated with the ocean carbon cycle in marine surface waters. The apparent resource partitioning of DOM by different bacterial species that was suggested by the data supports the significance of microbial community dynamics in the ocean's carbon cycle. The findings also underscore the importance of describing microbial synergistic interactions and population dynamics occurring on relatively short time-scales of hours to days.

Methods

Microcosm Setup and Biomass Sampling. Seawater for microcosm incubation experiments was collected (23°12.88′ N, 159°8.17′ W) from 75-m depth, predawn, on August 16, 2007, during the Center for Microbial Oceanography: Research and Education (C-MORE) BLOOMER Cruise. See *SI Appendix* for further details on the seawater collection and microcosm preparation.

HMW DOM Preparation. Surface seawater obtained from the uncontaminated underway system of the R/V *Kilo Moana* was filtered to remove microbes and small particles using a clean (10% HCl overnight soak), 0.2-μm Whatman Polycap TC polyether sulfone capsule filter. HMWDOM was concentrated using a custom-built ultrafiltration apparatus equipped with a stainless-steel membrane housing and centripetal pump along with a fluorinated high-density polyethylene reservoir. The system was plumbed with Teflon tubing and PVDF valves, and fitted with a dual thin-film ultrafiltration membrane element (Separation Engineering). The membrane has a 1-nm pore size that nominally retains organic matter of a molecular weight greater than 1,000 Da (>98% rejection of vitamin B_{12}). Membranes were precleaned with 0.01 mol L^{-1} hydrochloric acid (overnight wash) and 0.01 mol L^{-1} sodium hydroxide (overnight wash), and rinsed with copious amounts of distilled water until the pH returned to neutral. Membranes were flushed with 100 L of seawater for 45 min just before sample collection. Surface seawater (2,000 L) was concentrated 100-fold over a period of 24 h. Samples were taken for DOC quantification from the inflow and permeate during ultrafiltration, and of the concentrate upon completion. A 2-L subsample of the concentrate was prefiltered using a 0.2-μm Polycap TC filter (Whatman) before filtration through a prerinsed 30-kDa Ultracel regenerated cellulose membrane loaded in a high-output stirred cell (Millipore) to remove viral particles.

Dissolved Organic Carbon. DOC samples of 30 mL were transferred into combusted (450 °C for 8 h) glass vials and acidified with 150 mL of a 25% phosphoric acid solution before sealing with acid-washed Teflon septa and storage at 4 °C until processing. Analysis was performed using the high-temperature combustion method on a Shimadzu TOC-VCSH with platinized alumina catalyst. Sample concentrations were determined alongside potassium hydrogen phthalate standards and consensus reference materials (CRM) provided by the DOC-CRM program (http://www.rsmas.miami.edu/groups/biogeochem/CRM.html).

Flow Cytometry and Cell Sorting. At each time point, 1 mL of seawater was preserved with 0.125% glutaraldehyde (final concentration), frozen in liquid nitrogen, and stored at –80 °C for subsequent flow cytometric analysis and cell sorting using an Influx (Becton Dickinson). Before counting and sorting, samples were stained with SYBR Green (Invitrogen) for 15 min, and DNA-containing cells were identified based on fluorescence and scatter signals (43). See *SI Appendix* for further details on cell sorting and rRNA amplicon sequencing from the sorted population.

RNA Amplification and cDNA Synthesis. Metatranscriptome analyses were performed as previously described (44) with minor modifications. Briefly, 100 ng of total RNA was amplified using MessageAmp II (Ambion) following the manufacturer's instructions and substituting the T7-BpmI-$(dT)_{16}$VN oligo (44) in place of that supplied with the kit. Amplified RNA was then reverse transcribed into cDNA using SuperScript Double-Stranded cDNA Synthesis kit (Invitrogen) and random hexamer priming. Last, the cDNA was digested with BpmI and used for pyrosequencing. See *SI Appendix* for further details on pyrosequencing.

Bioinformatic Analyses. Full-length SSU rDNA amplicon sequences from flow-sorted cells were classified using both the Greengenes (45) NAST aligner and the Ribosomal Database Project (RDP) naïve Bayesian classifier (46). Resulting alignments were compared with the SILVA (47) databases using ARB (48). RDP classifier results were compared also with type strains using tools available at the RDP (49) and Interactive Tree of Life web sites (50).

cDNA datasets were parsed to separate rRNA sequences from the remaining non-rRNA sequences. rRNA sequences were identified as previously described (44) using a bit-score cutoff of 40 for BLASTN (32) searches against a custom 5S, SSU, 18S, 23S, and 28S rRNA databases. Non-rRNA sequences were compared with NCBI-nr, KEGG, and GOS protein clusters databases using BLASTX (32) for functional gene analyses as previously described (29, 44). See *SI Appendix* for further details.

Statistical Analyses. Statistical analyses were conducted on KEGG ortholog groups using the packages DegSeq (51) and ShotgunFunctionalizeR (52) in the R Statistical Package (53). In all statistical analyses, we assumed that the data (counts for a particular KEGG ortholog group) followed a Poisson sampling distribution. Analyses were conducted at the individual gene level as well as at the pathway level. See *SI Appendix* for further details on statistical analyses.

Accession Numbers. All 454 FLX pyrosequencing .sff files have been deposited in the GenBank database under accession no. SRA020733.11. Full-length SSU SSU rRNA sequences obtained from flow-sorted cells have been deposited to the GenBank/EMBL/DDJB databases under accession nos. HQ012268-HQ012278.

ACKNOWLEDGMENTS. We thank the captain and crew of the R/V *Kilo Moana* for facilitating sample collection, Chief Scientist Ricardo Letelier and all participants of the C-MORE BLOOMER cruise for help and encouragement, and Rachel Barry for pyrosequence library production and sequencing. This work was supported by the Gordon and Betty Moore Foundation (E.F.D., S.W.C., and D.J.R.), the Office of Science–Biological and Environmental Research, US Department of Energy (E.F.D and S.W.C), the National Science Foundation (D.J.R.), and National Science Foundation Science and Technology Center Award EF0424599 (to E.F.D. and S.W.C.). This article is a contribution from the National Science Foundation Science and Technology Center for Microbial Oceanography: Research and Education (C-MORE).

1. Hedges JI (1992) Global biogeochemical cycles: Progress and problems. *Mar Chem* 39: 67–93.
2. Ogawa H, Amagai Y, Koike I, Kaiser K, Benner R (2001) Production of refractory dissolved organic matter by bacteria. *Science* 292:917–920.
3. Pomeroy LR (1974) Oceans food web, a changing paradigm. *Bioscience* 24:499–504.
4. Azam F, et al. (1983) The ecological role of water-column microbes in the sea. *Mar Ecol Prog Ser* 10:257–263.
5. Azam F (1998) Microbial control of oceanic carbon flux: The plot thickens. *Science* 280: 694–696.
6. Ducklow H (1999) The bacterial component of the oceanic euphotic zone. *FEMS Microbiol Ecol* 30:1–10.
7. Benner R, Pakulski JD, McCarthy M, Hedges JI, Hatcher PG (1992) Bulk chemical characteristics of dissolved organic matter in the ocean. *Science* 255:1561–1564.
8. Church MJ, Ducklow HW, Karl DM (2002) Multiyear increases in dissolved organic matter inventories at station ALOHA in the North Pacific Subtropical Gyre. *Limnol Oceanogr* 47:1–10.
9. Carlson C, et al. (2002) Effect of nutrient amendments on bacterioplankton production, community structure, and DOC utilization in the northwestern Sargasso Sea. *Aquat Microb Ecol* 30:19–36.
10. Carlson C, et al. (2004) Interactions among dissolved organic carbon, microbial processes, and community structure in the mesopelagic zone of the northwestern Sargasso Sea. *Limnol Oceanogr* 49:1073–1083.
11. Pinhassi J, Berman T (2003) Differential growth response of colony-forming alpha- and gamma-proteobacteria in dilution culture and nutrient addition experiments from Lake Kinneret (Israel), the Eastern Mediterranean Sea, and the Gulf of Eilat. *Appl Environ Microbiol* 69:199–211.
12. Pinhassi J, et al. (2004) Changes in bacterioplankton composition under different phytoplankton regimens. *Appl Environ Microbiol* 70:6753–6766.
13. Schafer H, et al. (2001) Microbial community dynamics in Mediterranean nutrient-enriched seawater mesocosms: Changes in the genetic diversity of bacterial populations. *FEMS Microbiol Ecol* 34:243–253.
14. Allers E, et al. (2007) Response of Alteromonadaceae and Rhodobacteriaceae to glucose and phosphorus manipulation in marine mesocosms. *Environ Microbiol* 9:2417–2429.
15. Cecilia A, Jakob P (2006) *Roseobacter* and SAR11 dominate microbial glucose uptake in coastal North Sea waters. *Environ Microbiol* 8:2022–2030.
16. Hansell DA, Carlson CA (2001) Biogeochemistry of total organic carbon and nitrogen in the Sargasso Sea: Control by convective overturn. *Deep Sea Res Part II Top Stud Oceanogr* 48:1649–1667.
17. Morris RM, et al. (2005) Temporal and spatial response of bacterioplankton lineages to annual convective overturn at the Bermuda Atlantic Time-Series Study site. *Limnol Oceanogr* 50:1687–1696.
18. Quince C, et al. (2009) Accurate determination of microbial diversity from 454 pyrosequencing data. *Nat Methods* 6:639–641.
19. Kunin V, Engelbrektson A, Ochman H, Hugenholtz P (2010) Wrinkles in the rare biosphere: Pyrosequencing errors can lead to artificial inflation of diversity estimates. *Environ Microbiol* 12:118–123.
20. Turnbaugh PJ, et al. (2010) Organismal, genetic, and transcriptional variation in the deeply sequenced gut microbiomes of identical twins. *Proc Natl Acad Sci USA* 107: 7503–7508.
21. Stewart FJ, Ottesen EA, Delong EF (2010) Development and quantitative analyses of a universal rRNA-subtraction protocol for microbial metatranscriptomics. *ISME J* 4: 896–907.
22. Vila-Costa M (2010) Transcriptomic analysis of a marine bacterial community enriched with dimethylsulfoniopropionate. *ISME J*, 10.1038/ismej.2010.62.
23. Gilbert JA, et al. (2008) Detection of large numbers of novel sequences in the metatranscriptomes of complex marine microbial communities. *PLoS ONE* 3:e3042.
24. Ivanova EP, Flavier S, Christen R (2004) Phylogenetic relationships among marine Alteromonas-like proteobacteria: Emended description of the family Alteromonadaceae and proposal of Pseudoalteromonadaceae fam. nov., Colwelliaceae fam. nov., Shewanellaceae fam. nov., Moritellaceae fam. nov., Ferrimonadaceae fam. nov., Idiomarinaceae fam. nov. and Psychromonadaceae fam. nov. *Int J Syst Evol Microbiol* 54:1773–1788.
25. Hou S, et al. (2004) Genome sequence of the deep-sea gamma-proteobacterium Idiomarina loihiensis reveals amino acid fermentation as a source of carbon and energy. *Proc Natl Acad Sci USA* 101:18036–18041.
26. Ivars-Martinez E, et al. (2008) Comparative genomics of two ecotypes of the marine planktonic copiotroph Alteromonas macleodii suggests alternative lifestyles associated with different kinds of particulate organic matter. *ISME J* 2:1194–1212.
27. Giovannoni SJ, et al. (2005) Genome streamlining in a cosmopolitan oceanic bacterium. *Science* 309:1242–1245.
28. Tripp HJ, et al. (2008) SAR11 marine bacteria require exogenous reduced sulphur for growth. *Nature* 452:741–744.
29. Shi Y, Tyson GW, DeLong EF (2009) Metatranscriptomics reveals unique microbial small RNAs in the ocean's water column. *Nature* 459:266–269.
30. Kanehisa M, et al. (2008) KEGG for linking genomes to life and the environment. *Nucleic Acids Res* 36:D480–D484.
31. Yooseph S, et al. (2007) The Sorcerer II Global Ocean Sampling expedition: Expanding the universe of protein families. *PLoS Biol* 5:e16.
32. Altschul SF, et al. (1997) Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. *Nucleic Acids Res* 25:3389–3402.
33. Zimmer DP, et al. (2000) Nitrogen regulatory protein C-controlled genes of Escherichia coli: Scavenging as a defense against nitrogen limitation. *Proc Natl Acad Sci USA* 97:14674–14679.
34. Mannino A, Harvey HR (1999) Lipid composition in particulate and dissolved organic matter in the Delaware Estuary: Sources and diagenetic patterns. *Geochim Cosmochim Acta* 63:2219–2235.
35. Repeta DJ, Aluwihare LI (2006) High molecular weight dissolved organic carbon cycling as determined by natural abundance radiocarbon measurements of neutral sugars. *Limnol Oceanogr* 51:1045–1053.
36. Lou Y-W, Friedrichs MAM, Doney SC, Church MJ, Ducklow HW (2010) Oceanic heterotrophic bacterial nutrition by semilabile DOM as revealed by data assimilative modeling. *Aquat Microb Ecol* 60:273–287.
37. Panagiotopoulos C, Repeta DJ, Johnson CG (2007) Characterization of methyl sugars, 3-deoxysugars and methyl deoxysugars in marine high molecular weight dissolved organic matter. *Org Geochem* 38:884–896.
38. Quan TM, Repeta DJ (2007) Characterization of high molecular weight dissolved organic carbon using periodate over-oxidation. *Mar Chem* 105:183–193.
39. Lidstrom ME (2006) Aerobic methylotrophic prokaryotes. *The Prokaryotes*, ed Dworkin M (Springer, New York), 3rd Ed, pp 618–634.
40. Neufeld JD, Boden R, Moussard H, Schafer H, Murrell JC (2008) Substrate-specific clades of active marine methylotrophs associated with a phytoplankton bloom in a temperate coastal environment. *Appl Environ Microbiol* 74:7321–7328.
41. Neufeld JD, Chen Y, Dumont MG, Murrell JC (2008) Marine methylotrophs revealed by stable-isotope probing, multiple displacement amplification and metagenomics. *Environ Microbiol* 10:1526–1535.
42. Neufeld JD, et al. (2007) Stable-isotope probing implicates Methylophaga spp and novel Gammaproteobacteria in marine methanol and methylamine metabolism. *ISME J* 1:480–491.
43. Marie D, Partensky F, Jacquet S, Vaulot D (1997) Enumeration and cell cycle analysis of natural populations of marine picoplankton by flow cytometry using the nucleic acid stain SYBR Green I. *Appl Environ Microbiol* 63:186–193.
44. Frias-Lopez J, et al. (2008) Microbial community gene expression in ocean surface waters. *Proc Natl Acad Sci USA* 105:3805–3810.
45. DeSantis TZ, et al. (2006) Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. *Appl Environ Microbiol* 72:5069–5072.
46. Wang Q, Garrity GM, Tiedje JM, Cole JR (2007) Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. *Appl Environ Microbiol* 73:5261–5267.
47. Pruesse E, et al. (2007) SILVA: A comprehensive online resource for quality checked and aligned ribosomal RNA sequence data compatible with ARB. *Nucleic Acids Res* 35: 7188–7196.
48. Ludwig W, et al. (2004) ARB: A software environment for sequence data. *Nucleic Acids Res* 32:1363–1371.
49. Cole JR, et al. (2009) The Ribosomal Database Project: Improved alignments and new tools for rRNA analysis. *Nucleic Acids Res* 37:D141–D145.
50. Letunic I, Bork P (2007) Interactive Tree of Life (iTOL): An online tool for phylogenetic tree display and annotation. *Bioinformatics* 23:127–128.
51. Wang L, Feng Z, Wang X, Wang X, Zhang X (2010) DEGseq: An R package for identifying differentially expressed genes from RNA-seq data. *Bioinformatics* 26: 136–138.
52. Kristiansson E, Hugenholtz P, Dalevi D (2009) ShotgunFunctionalizeR: An R-package for functional comparisons of metagenomes. *Bioinformatics* 25:2737–2738.
53. R Development Core Team (2010) *R: A language and environment for statistical computing, version 2.11.1* (R Foundation for Statistical Computing, Vienna). Available at http://www.r-project.org/.

SUPPORTING INFORMATION

SUPPORTING TEXT

METHODS

Microcosm Setup and Biomass Sampling. Seawater for microcosm incubation experiments was collected at 23°12.88' N, 159°8.17'W, from 75-m depth, pre-dawn on August 16, 2007 during the CMORE Bloomer Cruise. Hydrocasts for sampling were conducted using a conductivity-temperature-depth (CTD) rosette sampler aboard the R/V Kilo Moana. Water was transferred to acid-washed, then sample-water rinsed 20L polycarbonate bottles. The incubator was a blue light type, which simulated the light levels at ~25-45m depth (roughly 14% surface irradiance). The carboys were wrapped 4x in black fiberglass screen, to further decrease the light levels inside the carboy to 3% surface irradiance, the *in situ* light intensity at 75m. These bottles were incubated in deck-board incubators supplied with flow-through surface seawater to maintain near *in situ* temperatures (approximate 0.6°C temperature differential between 75m and sea surface during the course of experiment). 2L of HMW DOM concentrate was added to 18L source water for a total initial volume of 20L and final DOC concentration of 328 μM C, approximately 4x the ambient value of 82 μM C. Replicate control and HMW DOM amended microcosms were initiated at 05:45 local time with subsamples taken at 2, 6, 12, 19, and 27 hours post HMW DOM addition. At selected timepoints, microbial biomass from ~2L was rapidly collected for RNA samples by first pre-filtering through a 1.6μm glass fiber filter and then harvesting cells onto 0.2μm durapore (Millipore, Billerica MA). Filtration was limited to less than 10 minutes and then the filter was immediately placed into RNAlater (Applied Biosystems, Foster City CA) and frozen at -80°C. RNA

extraction, purification, and DNAse treatments performed as previously described(1). At both the beginning and the end of the experiment 10L was similarly sampled for DNA first by pre-filtration through a 1.6 μm glass fiber filter and then collected onto 0.2μm Sterivex (Millipore) filters. DNA extraction and purification performed as previously described(2).

Flow Cytometry and Cell Sorting. At each time point 1 ml of seawater was preserved with 0.125% glutaraldehyde (final conc.), frozen in liquid N^2, and stored at –80°C for subsequent flow cytometric analysis and cell sorting using an Influx (Becton Dickinson). Prior to counting and sorting, samples were stained with Sybr green (Invitrogen, Carslbad CA) for 15 min, and DNA-containing cells were identified based on fluorescence and scatter signals (3).

A population of large non-pigmented cells appearing in DOM-amended incubations was sorted for identification by 16S rRNA gene sequencing. Approximately 40,000 cells from the final time point sample were first sorted into clean sheath fluid, then re-sorted directly into eight PCR tubes. Two rounds of sorting helped eliminate co-transport of dissolved DNA and ensured that only the targeted cells were amplified(4). Amplifications of 16S rRNA genes from flow-sorted cells were performed with universal 6F and 1492R primers, and the resulting amplification products pooled. These pooled PCR products were cloned using a TOPO-TA kit (Invitrogen, Carslbad CA) and paired end reads sequenced using BigDye v3.1 chemistry on an ABI 3730 capillary sequencer (Appplied Biosystems, Foster City CA).

To prepare the Influx for clean sorting, fluid lines were flushed with 10% bleach for 20 min and rinsed with UV-treated MilliQ for 10min. Fluid lines where then dried by

pumping air through for 10min before leaving overnight. Sheath fluid (1% NaCl w/v), sample tubes, and the sheath tank were UV-treated for 90min then left overnight, then re-treated with UV for 5min the following morning.

RNA Amplification and cDNA Synthesis. Performed as previously described(1) with minor modifications. Briefly, 100 ng of total RNA was amplified using MessageAmp II (Ambion, Foster City CA) following the manufacturer's instructions and substituting the T7-BpmI-$(dT)_{16}$VN oligo(1) in place of that supplied with the kit. Amplified RNA was then reverse transcribed into cDNA using Superscript Double-Stranded cDNA Synthesis kit (Invitrogen) and random hexamer priming. Lastly the cDNA was digested with BpmI and utilized for pyrosequencing.

Pyrosequencing. For both DNA and cDNA libraries, 1μg of material was used for sequencing with a Roche FLX 454 sequencer yielding on average 251074 and 241462 reads per run, respectively. In general, two runs were combined for each library. cDNA sequence libraries were dominated by SSU SSU rRNA sequences which represented 93-95% of the total reads. Various commercial kits and enzymes are available to selectively remove or reduce the relative abundance of rRNAs in total RNA extracts, however these treatments have produced limited beneficial results in our hands. While increasing the proportion of reads from non rRNA molecules is desirable, the large number of reads obtained here remain useful for taxonomic classification of these microbial communities.

Bioinformatic Analyses. Full-length SSU rDNA sequences from flow-sorted cells were classified using both the Greengenes (5) NAST aligner and the Ribosomal Database Project (RDP) naïve Bayesian classifier (6). Resulting alignments were compared to the SILVA(7) databases using ARB (8). Additionally RDP classifier results were compared to type strains utilizing tools available at the RDP (9) and Interactive Tree of Life websites (10).

cDNA datasets were parsed to separate rRNA sequences from the remaining non-rRNA sequences. rRNA sequences were identified as previously described (1) utilizing a bit score cutoff of 40 for BLASTN (11) searches against a custom 5S, SSU, 18S, 23S, and 28S rRNA database.

MEGAN software (12) was employed for analyzing the taxonomic breakdown of non-rRNA cDNAs with bit scores > 40 within 10% of the top scoring hits. SSU rRNA pyrosequencing reads from both DNA and cDNA datasets using were also analyzed in MEGAN using these same settings in conjunction with specialized SSU databases developed by Urich *et al.* (13).

Non-rRNA sequences were compared to NCBI-nr, KEGG, and GOS protein clusters databases using BLASTX (11) for functional gene analyses. Top hits with bit scores > 40 were used to assign reads to individual proteins/peptides with the KEGG database results used primarily. Assignment of reads to individual KEGG ortholog groups allowed for enumeration and comparison of ortholog abundance between amended and control microcosms. KEGG ortholog abundance values were normalized by the number of reads in each respective library. In comparisons against the reference

genomes *Idiomarina loihiensis* and *Alteromonas macleodii Deep ecotype* DSM 17117, normalization to reference genome gene sizes was also performed.

Reference genome sequence comparisons. Non-rRNA reads at all time points were compared against a custom database of amino acid sequences compiled from publicly available microbial genomes (fully sequenced and draft genomes as of January 2009). Reads with top hits with bits scores > 50 were assigned to the corresponding genomes. To identify differentially expressed ORFs with statistical significance in a reference genome, the reference genome needs to be well represented in both control and treatment data sets. Therefore we used the genomes of, *Pelagibacter* strain HTCC 7211, *Prochlorococcus* strain AS9601 as reference genomes. Statistical analysis was performed with the R package DEGseq (14), under the following settings: FET (Fisher's Exact Test), q-value (a measure of significance in terms of false discovery rate) of 0.005.

Because of their consistently low representation in the control datasets and high abundance in the DOM amendment microscosm, pairwise statistical comparisons were not possible, so a different approach and separate analyses had to be used for *Idiomarina loihiensis* and *Alteromonas macleodii Deep ecotype* DSM 17117. Rank abundance tables (Table S1 and Table S2) of non-rRNA cDNAs with bit scores > 40 within 10% of the top scoring hits to reference genomes of *Idiomarina loihiensis* and *Alteromonas macleodii Deep ecotype* DSM 17117 were binned. The abundance per of each KEGG homolog, corrected for the specific gene size in the reference taxon, and normalized to the total

nucleotide count of KEGG homologs in each taxon. The normalized KEGG homolog abudances were then tabulated, ordered and compared manually. 2.

Small RNA analyses. Putative sRNA (psRNA) clusters were identified using a pipeline modified from an early version (15). Briefly, CD-HIT (16) was used to cluster cDNA sequences with the following parameters: -c 0.90 –n 7 -r 1 -g 1 -G 0 -aS 0.9 -p 1 -d 0 -b 10, which translates to clustering at 90% sequence identity over 90% of the length of the shorter read. A second iteration of clustering was performed at 85% sequence identity over 80% of the shorter length. The seed sequences of identified clusters were then extracted, subjected to an all-vs-all BLAST, and clustered with the self-clustering method described previously (15), using the cutoff of > 85% sequence identity, alignment length >100bp, and alignment start/stop position within 5bp to either end of the sequences (to avoid clustering two reads based on conserved regions or repeats in the middle part of the sequences). The purpose of doing iterative clustering with CD-HIT followed by the self-clustering method is 1) to reduce CPU time on large data sets (CD-HIT is ultra-fast by using short word filtering method whereas all-vs-all BLAST is computationally demanding), and 2) to allow the clustering of sequences that overlap at high sequence identity in the ends using the self-clustering method. Sequence clusters with > 100 reads were further examined to exclude apparent protein-coding clusters from further analyses. Characterization of the resulting psRNA clusters, including Rfam annotation, genomic context, etc., was performed as described previously (15).

Statistical Analyses: Statistical analyses were conducted on KEGG ortholog groups using the packages *ShotgunFunctionalizeR* (17) and *DegSeq* (18) in the R Statistical Package (19). In all statistical analyses, we assumed that the data (counts for a particular KEGG ortholog group) follow a Poisson sampling distribution. Analyses were conducted at the individual gene level as well as at the pathway level.

Poisson ANOVA was used to test for significant differences across treatments using modified functions in *ShotgunFunctionalizeR* (17). The modifications allow the use of an exposure term to properly scale the library sizes (N_i), where the size for library i is defined as the number of non-rRNA reads in the library (Table 1). The loglinear model is:

$$\ln\left(\frac{E[Y_{i,j}]}{N_i}\right) = \alpha_0 + \sum_{k=1}^{K} \alpha_{j,k} X_{j,k},$$

where X is a design matrix indicating whether ortholog j from library i belongs to treatment group k, and the estimated coefficients, α, include the intercept term (α_0, i.e., the grand mean) and the treatment effects ($\alpha_{j,k}$) of group k (e.g., HMW DOM addition or time in hours since the beginning of the experiment). Note that the coefficients are on the natural log scale, and since the libraries are scaled with the exposure term, the sum of the coefficients describing a treatment are interpreted (after exponentiating) as the proportion of (non-rRNA) genetic material in the population represented by ortholog j. Fisher's exact test was used for pairwise comparisons among genes. False discovery rates were controlled using the method of Storey (20), and reported q-values are calibrated for each table of comparisons.

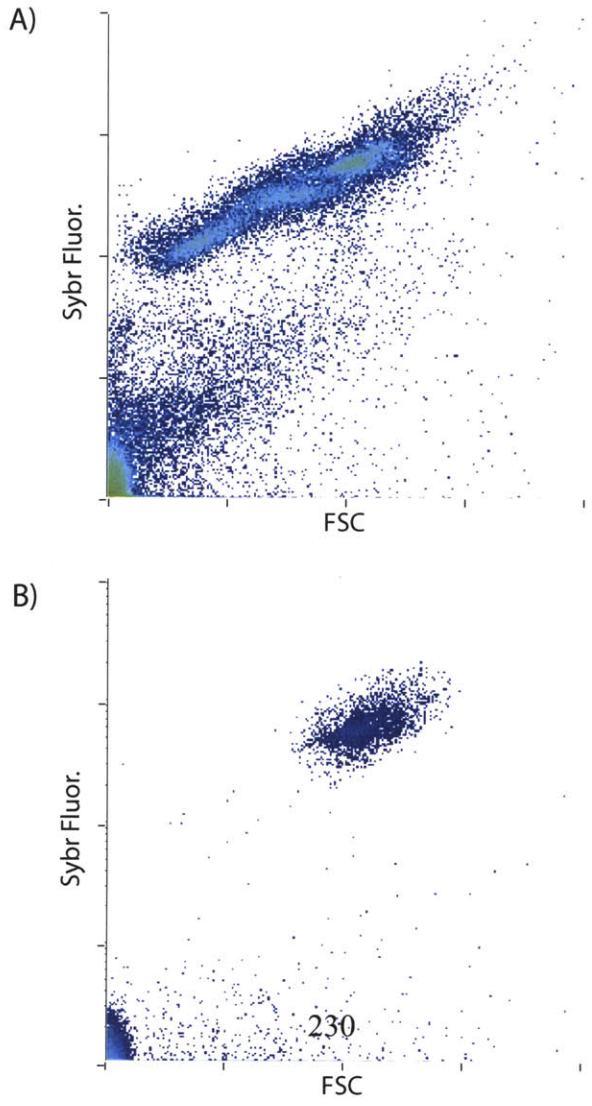

A)
Sybr Fluor.
FSC
B)
Sybr Fluor.
FSC

Figure S1. Flow cytometric scatter plots of SYBR-stained cells from (A) HMW DOM amended microbial community and (B) the population of cells resulting from flow cytometric sorting of the larger, higher DNA content cells present at the end of the experiment.

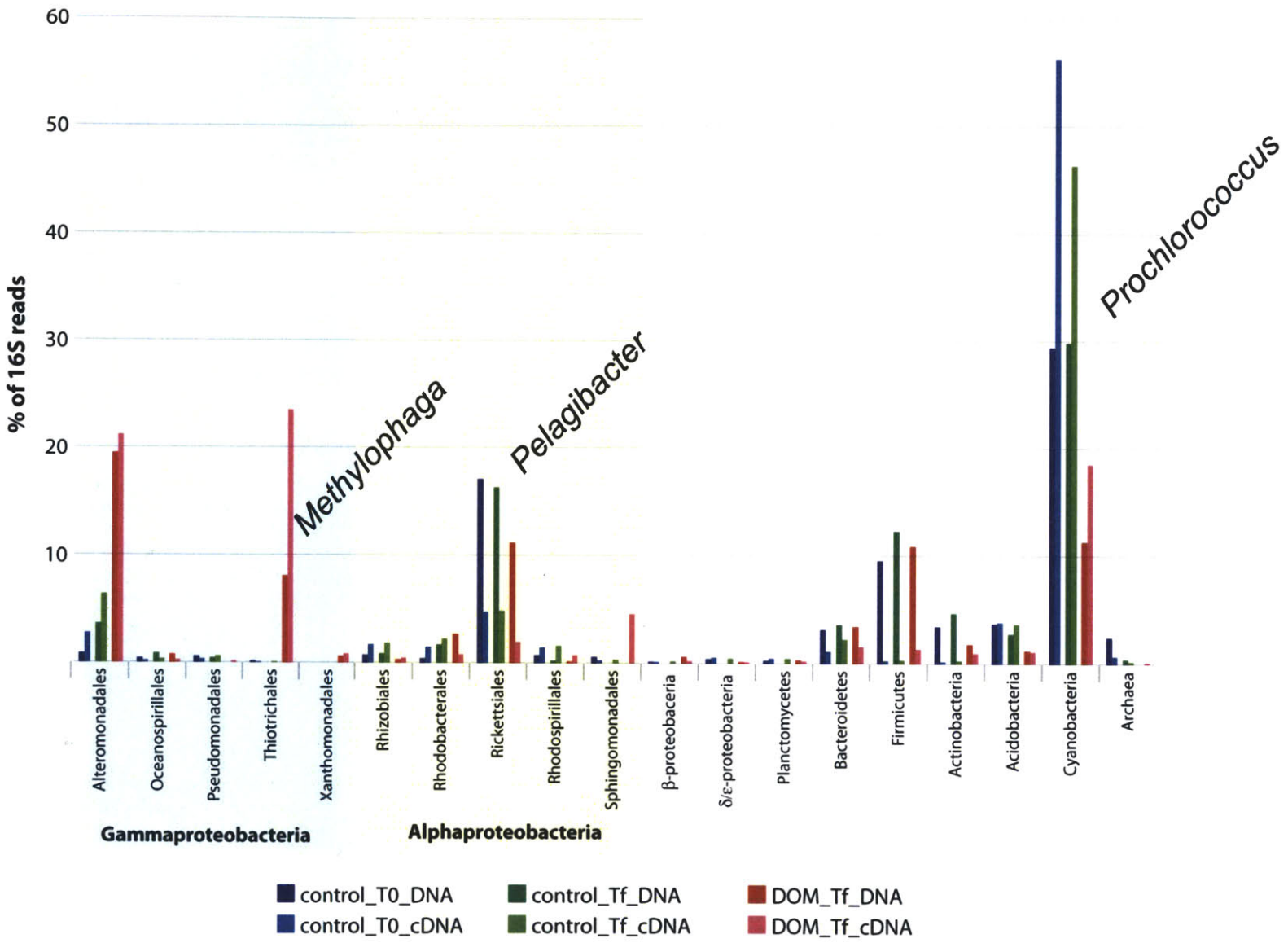

60
50
40
30
20
10
% of 16S reads
Methylophaga
Pelagibacter
Prochlorococcus
Alteromonadales
Oceanospirillales
Pseudomonadales
Thiotrichales
Xanthomonadales
Rhizobiales
Rhodobacterales
Rickettsiales
Rhodospirillales
Sphingomonadales
β-proteobacteria
δ/ε-proteobacteria
Planctomycetes
Bacteroidetes
Firmicutes
Actinobacteria
Acidobacteria
Cyanobacteria
Archaea
Gammaproteobacteria
Alphaproteobacteria
control_T0_DNA
control_T0_cDNA
control_Tf_DNA
control_Tf_cDNA
DOM_Tf_DNA
DOM_Tf_cDNA

Figure S2. Microbial community composition assessed by taxonomic classification of metagenomic and metatranscriptomic sequence reads. SSU rRNA reads were extracted from both DNA as well as cDNA datasets, and classified according to phylogenetic groups (see Supporting Information Methods, above).

Fig S3 Differentially expressed ORFs in *Pelagibacter* HTCC 7211

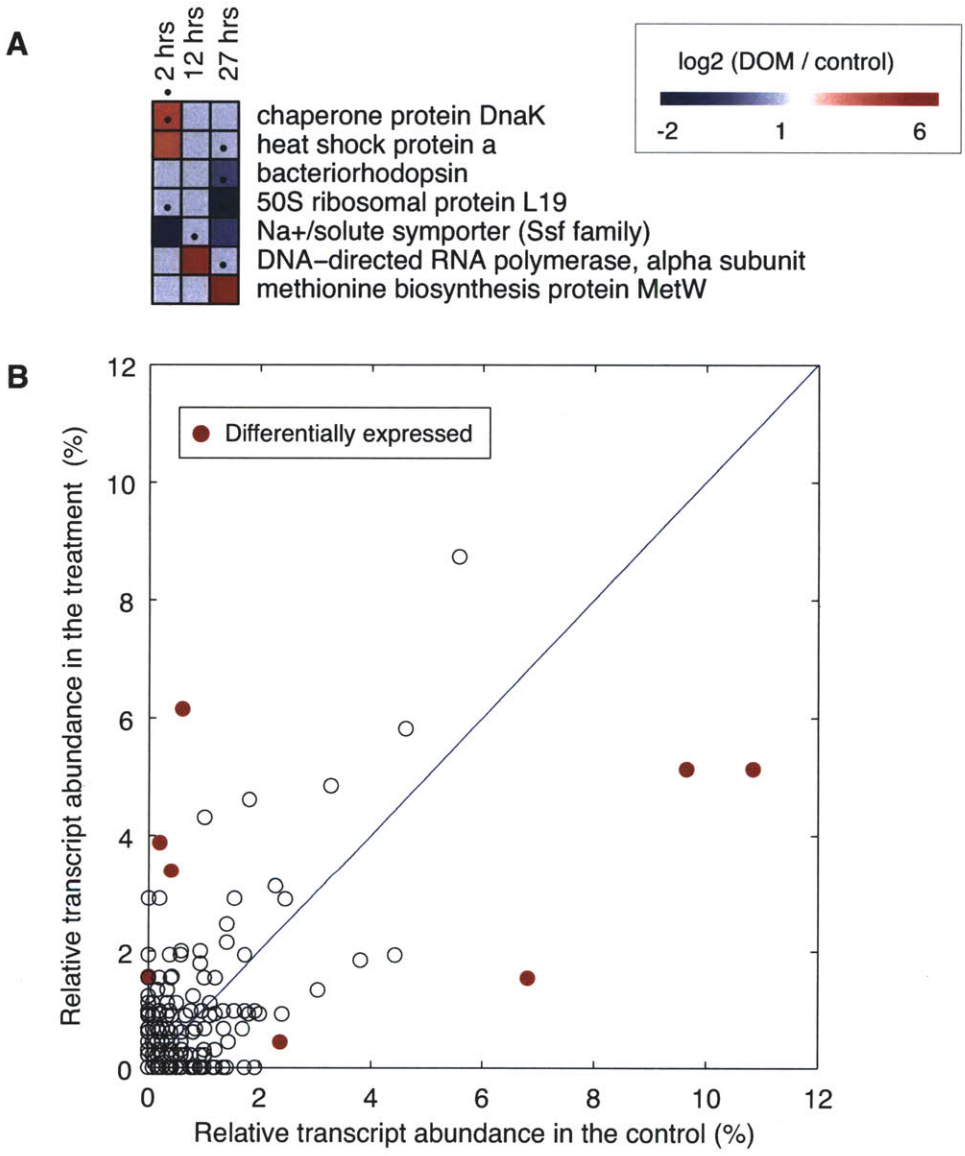
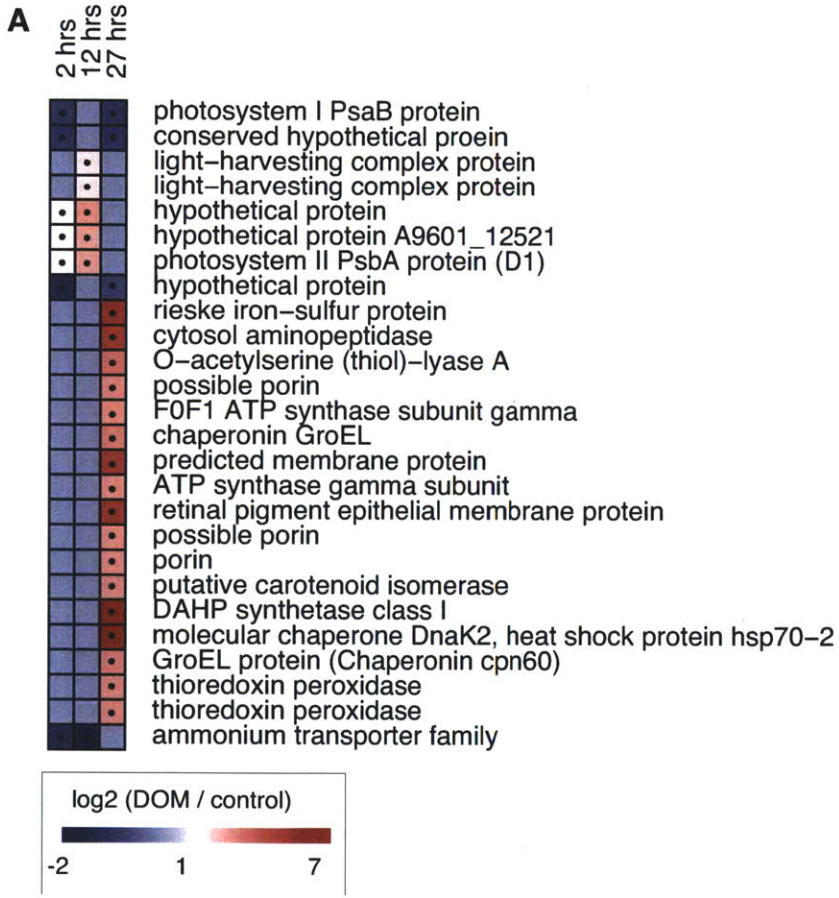


Figure S3. Differentially expressed ORFs in the reference genome of *Pelagibacter* strain HTCC 7211. (A) Heatmap of ORFs with statistically significant (q-value < 0.005) differential abundance at any of the three time points. Black dots indicate the time point the ORF was differentially expressed in the treatment. (B) Percentage of detected ORFs in the treatment and control at each of the three time points, relative to all reads assigned to *Pelagibacter* strain HTCC 7211 at the corresponding time point. Red dots represent ORFs whose relative abundance difference was considered statistically significan

Fig S4

Differentially expressed ORFs in *Prochlorococcus* AS9601



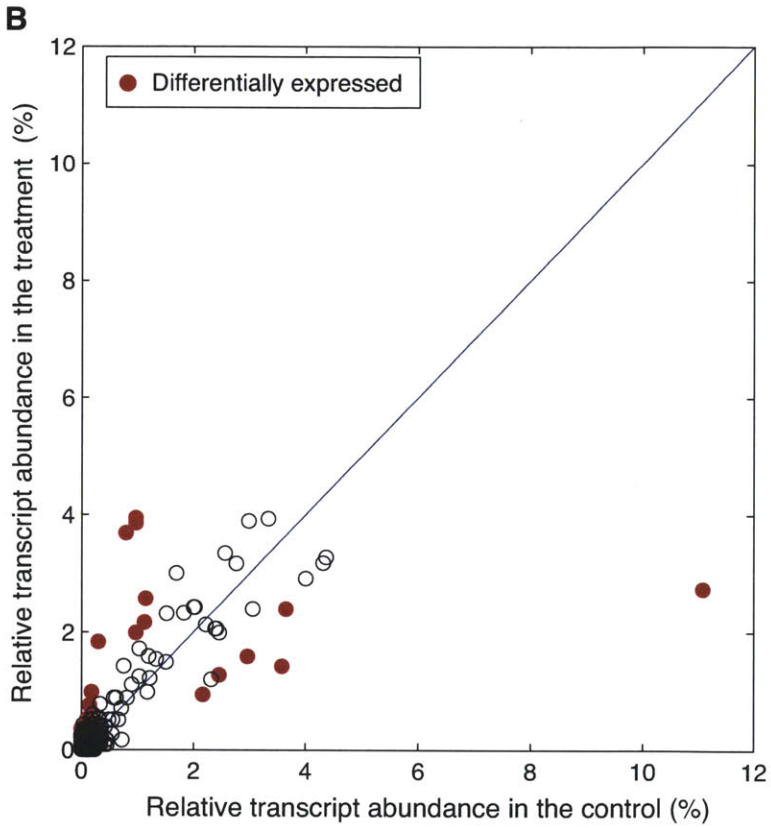
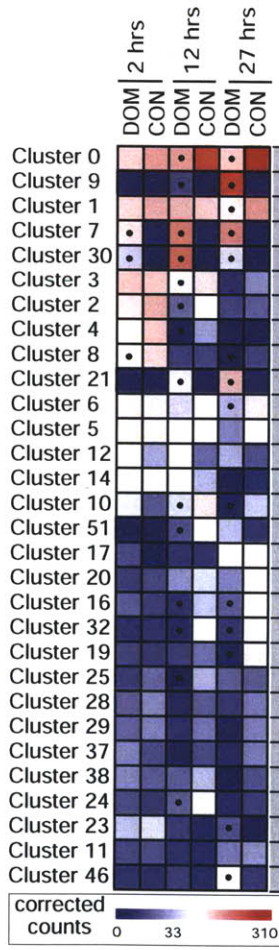


Figure S4. Differentially expressed ORFs in the reference genome of *Prochlorococcus* strain AS9601. (A) Heatmap of ORFs with statistically significant (q-value < 0.005) differential abundance at any of the three time points. Black dots indicate the time point the ORF was differentially expressed in the treatment. (B) Percentage of detected ORFs in the treatment and control at each of the three time points, relative to all reads assigned to *Prochlorococcus* strain AS9601 at the corresponding time point. Red dots represent ORFs whose relative abundance difference was considered statistically significant.



	Rfam annotation[a]	H179 psRNA groups[b]	Flanking ORF annotation	Putative taxonomic assignment[c]
Cluster 0	NA	NA	Integral membrane protein (DMT superfamily)	Bacteria; Chlamydiae/Verrucomicrobia group
Cluster 9	tmRNA	NA	SsrA-binding protein	Bacteria ;Proteobacteria ;Gammaproteobacteria;Methylophaga
Cluster 1	RNaseP_bact_a	Group_9	dsRNA-specific ribonuclease	Bacteria; Cyanobacteria; Prochlorales
Cluster 7	tmRNA	NA	SsrA-binding protein	Bacteria; Gammaproteobacteria; Idiomarinaceae
Cluster 30	tmRNA	NA	2-polyprenylphenol hydroxylase	Bacteria; Gammaproteobacteria; Oceanospirillaceae
Cluster 3	NA	NA	fumarate hydratase class II	Bacteria; environmental samples
Cluster 2	NA	NA	fumarate hydratase, class II	Eukaryota; Viridiplantae; Chlorophyta
Cluster 4	NA	Group_16	fumarate hydratase, class II	Bacteria; Spirochaetes
Cluster 8	NA	NA	fumarate hydratase, class II	Bacteria; environmental samples
Cluster 21	tmRNA	NA	SsrA-binding protein	Bacteria; Gammaproteobacteria; Alteromonadaceae
Cluster 6	NA	NA	flavoprotein-ubiquinone oxidoreductase	Bacteria; Proteobacteria; Gammaproteobacteria; Alteromonadales; Alteromonadaceae
Cluster 5	NA	Group_28	short-chain dehydrogenase	Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales
Cluster 12	NA	NA	pyruvate kinase	Bacteria; environmental samples
Cluster 14	NA	NA	fumarate hydratase, class II	Bacteria; environmental samples
Cluster 10	NA	Group_58	Translation initiation factor 3 (IF-3)	Bacteria; Gammaproteobacteria; Idiomarinaceae
Cluster 51	NA	Group_2	hypothetical protein	Bacteria; Cyanobacteria; Prochlorales
Cluster 17	NA	NA	50S ribosomal protein L19	Bacteria; Bacteroidetes; Flavobacteria; Flavobacteriales
Cluster 20	NA	NA	NA	NA
Cluster 16	NA	NA	endo-1,4-beta-xylanase	Bacteria; Planctomycetes; Planctomycetacia; Planctomycetales
Cluster 32	NA	NA	NA	NA
Cluster 19	NA	NA	NA	NA
Cluster 25	NA	NA	NA	NA
Cluster 28	NA	NA	tryptophanyl-tRNA	Bacteria; Proteobacteria; Deltaproteobacteria; Desulfovibrionales
Cluster 29	NA	Group_5	hypothetical protein	Bacteria; Proteobacteria; Alphaproteobacteria/Gammaproteobacteria
Cluster 37	NA	NA	hydroxymethylglutaryl-CoA	Bacteria; environmental samples
Cluster 38	tmRNA	NA	hypothetical protein	Bacteria; Bacteroidetes; Flavobacteria; Flavobacteriales
Cluster 24	NA	NA	NA	NA
Cluster 23	NA	NA	NA	NA
Cluster 11	NA	NA	putative neutral invertase-like protein	Bacteria; Cyanobacteria; Prochlorales
Cluster 46	NA	NA	AsnC family transcriptional regulator	Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales

[a] Rfam database 10.0 (http://rfam.sanger.ac.uk/) was used as reference.

[b] Sequence similarity to previously identified psRNA groups (Shi, Y., Tyson, G. W. & DeLong, E. F. Metatranscriptomics reveals unique microbial small RNAs in the ocean's water column. Nature 459, 266–269) were performed with CD-HIT.

[c] Putative taxonomic assignments were based on BLASTp of flanking ORFs against NR.

Figure S5. Relative abundance of putative sRNA (psRNA) clusters identified in the treatment and control data sets, normalized to the sum of identified psRNA reads for each data set. The thirty clusters contained > 100 cDNA reads and were identified using a pipeline modified from an early version (see Materials and Methods). Black dots indicate the time point at which the ORF was differentially expressed in the treatment compared to the control (q-value < 0.005). Rfam annotation, homology to previously identified psRNA groups, annotations of flanking ORFs, and putative taxonomic assignment were listed when possible.

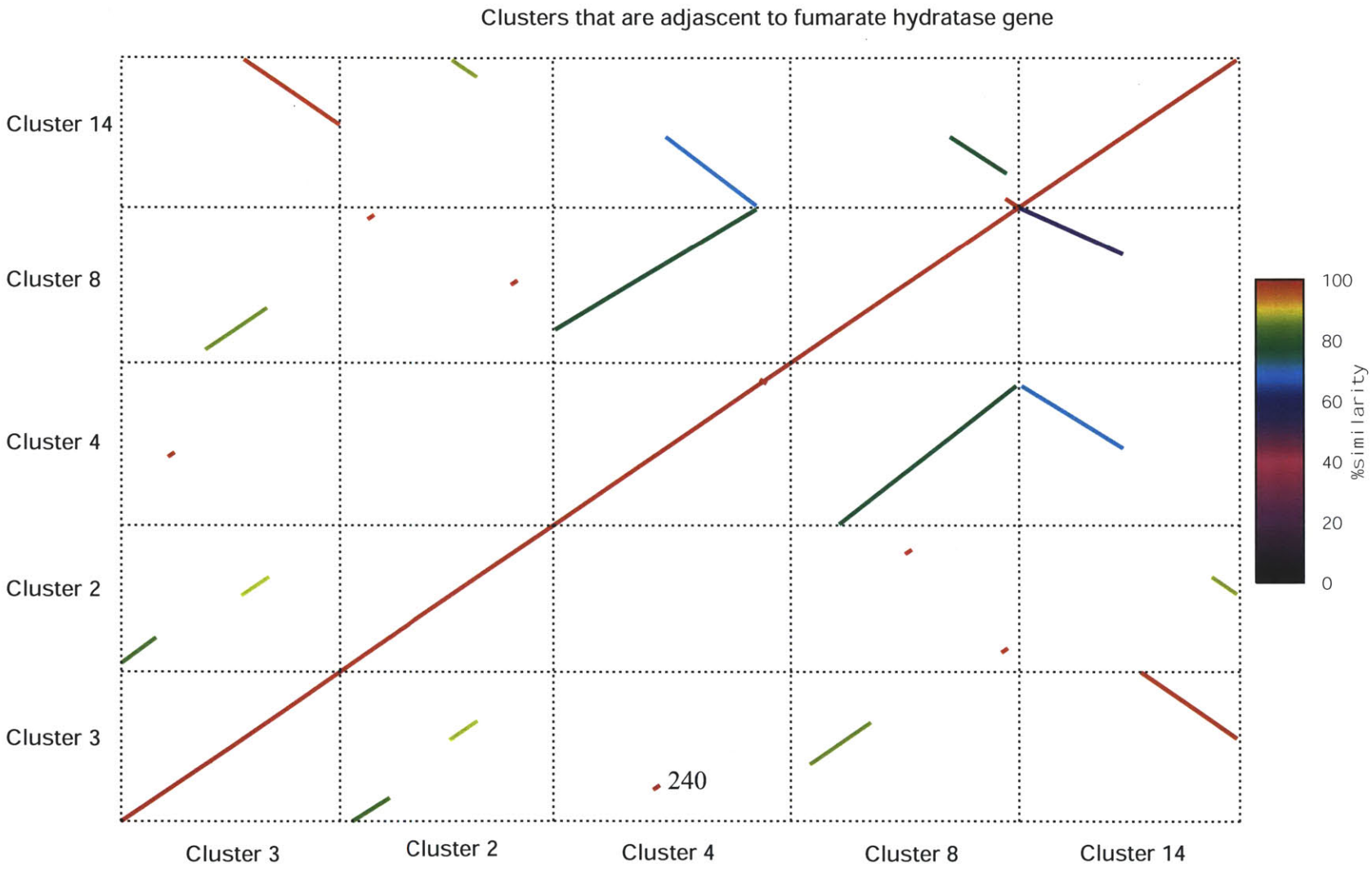

Clusters that are adjascent to fumarate hydratase gene
Cluster 14
Cluster 8
Cluster 4
Cluster 2
Cluster 3
240
Cluster 3
Cluster 2
Cluster 4
Cluster 8
Cluster 14
100
80
60
40
20
0
%similarity

Figure S6. Pair-wise alignment of psRNA clusters that were found adjacent to a gene encoding fumarate hydratase. The alignment was performed with NUCmer, part of the MUMmer 3.20 package (see Materials and Methods).

Table S1. Idiomarinaceae specific KEGG orthologues in control and treatment cDNAs

KO	Definition	con_2 hrs	con_1 2hrs	con_2 7hrs	DOM_ 2hrs	DOM_ 12hrs	DOM_ 27hrs
K00265	glutamate synthase (NADPH/NADH) large chain [EC:1.4.1.13 1.4.1.14]	0.00	0.00	0.00	6.34	2.12	0.68
K07486	transposase [NA]	0.00	0.00	0.00	4.99	5.35	3.69
K07576	metallo-beta-lactamase family protein [NA]	0.00	0.00	0.00	4.84	0.00	0.23
K07662	two-component system, OmpR family, response regulator CpxR [NA]	0.00	0.00	0.00	4.50	1.85	0.68
K02014	iron complex outermembrane recepter protein [NA]	0.11	0.43	0.15	4.48	1.14	1.26
K00266	glutamate synthase (NADPH/NADH) small chain [EC:1.4.1.13 1.4.1.14] [EC:1.	0.00	0.00	0.00	4.15	1.34	0.10
K01046	triacylglycerol lipase [EC:3.1.1.3] [EC:3.1.1.3]	0.00	0.00	0.00	2.94	0.19	0.20
K02406	flagellin [NA]	0.00	0.00	0.00	2.73	0.81	0.90
K00430	peroxidase [EC:1.11.1.7] [EC:1.11.1.7]	0.00	0.00	0.00	2.70	0.59	0.43
K02556	chemotaxis protein MotA [NA]	0.00	0.00	0.00	2.65	0.75	0.37
K01423	peptidase, M28 (aminopeptidase S) family (EC:3.4.-.-)	0.00	0.00	0.72	2.64	1.38	1.42
K03406	methyl-accepting chemotaxis protein [NA]	0.29	0.14	0.00	2.60	0.23	0.90
K00242	succinate dehydrogenase hydrophobic membrane anchor protein [EC:1.3.99.1	0.00	0.00	0.00	2.36	1.10	0.20
K02392	flagellar basal-body rod protein FlgG [NA]	0.30	0.00	0.00	2.33	0.00	0.00
K07659	two-component system, OmpR family, phosphate regulon response regulator (	0.00	0.00	0.00	2.25	0.53	0.10
K06445	acyl-CoA dehydrogenase [EC:1.3.99.-] [EC:1.3.99.-]	0.00	0.00	0.00	2.16	2.10	1.03
K07566	putative translation factor [NA]	0.00	0.00	0.00	2.07	0.00	0.48
K01638	malate synthase [EC:2.3.3.9] [EC:2.3.3.9]	0.00	0.00	0.10	2.06	1.45	0.09
K01637	isocitrate lyase [EC:4.1.3.1]	0.18	0.17	0.00	1.98	1.57	0.00
K03286	OmpA-OmpF porin, OOP family [NA]	0.00	0.00	0.00	1.73	0.20	0.82
K04085	tRNA 2-thiouridine synthesizing protein A [EC:2.8.1.-] [EC:2.8.1.-]	0.00	0.00	0.00	1.71	0.00	0.00
K01681	aconitate hydratase 1 [EC:4.2.1.3] [EC:4.2.1.3]	0.00	0.00	0.06	1.70	0.51	0.11
K02404	flagellar biosynthesis protein FlhF [NA]	0.00	0.00	0.00	1.68	0.29	0.05
K00799	glutathione S-transferase [EC:2.5.1.18] [EC:2.5.1.18]	0.00	0.00	0.00	1.57	0.88	0.32
K02387	flagellar basal-body rod protein FlgB [NA]	0.00	0.00	0.00	1.51	0.00	0.00
K00632	acetyl-CoA acyltransferase [EC:2.3.1.16] [EC:2.3.1.16]	0.19	0.00	0.00	1.48	1.24	0.45
K03111	single-strand DNA-binding protein [NA]	0.00	0.00	0.00	1.46	0.00	0.00
K02422	flagellar protein FliS [NA]	0.00	0.00	0.00	1.45	0.00	1.34
K04063	osmotically inducible protein OsmC [NA]	0.00	0.00	0.00	1.45	0.00	0.34
K01740	O-acetylhomoserine (thiol)-lyase [EC:2.5.1.49] [EC:2.5.1.49]	0.00	0.00	0.00	1.45	0.15	0.00
K07400	thioredoxin-like protein [NA]	0.00	0.00	0.00	1.42	0.89	0.00
K00413	ubiquinol-cytochrome c reductase cytochrome c1 subunit [EC:1.10.2.2] [EC:1.	0.00	0.00	0.00	1.39	0.26	0.19
K02391	flagellar basal-body rod protein FlgF [NA]	0.00	0.31	0.00	1.37	0.00	0.00
K02390	flagellar hook protein FlgE [NA]	0.00	0.00	0.00	1.36	0.28	0.00
K02895	large subunit ribosomal protein L24 [NA]	0.00	0.00	0.00	1.30	0.62	2.03
K03586	cell division protein FtsL [NA]	0.00	0.00	0.00	1.29	0.00	0.00
K01572	oxaloacetate decarboxylase, beta subunit [EC:4.1.1.3] [EC:4.1.1.3]	0.00	0.00	0.00	1.26	1.02	0.31
K02398	negative regulator of flagellin synthesis FlgM [NA]	0.00	0.00	0.00	1.26	0.00	0.21
K03408	purine-binding chemotaxis protein CheW [NA]	0.00	0.00	0.00	1.20	0.00	0.21
K03307	solute:Na+ symporter, SSS family [NA]	0.00	0.00	0.00	1.20	0.37	0.23
K07479	putative DNA topoisomerase [NA]	0.00	0.00	0.00	1.19	0.00	0.00
K01467	beta-lactamase [EC:3.5.2.6] [EC:3.5.2.6]	0.00	0.00	0.00	1.18	0.37	0.27
K04562	flagellar biosynthesis protein FlhG [NA]	0.00	0.00	0.00	1.17	0.22	0.16
K03117	sec-independent protein translocase protein TatB [NA]	0.00	0.00	0.00	1.16	0.00	0.00
K00945	cytidylateinase [EC:2.7.4.14] [EC:2.7.4.14]	0.00	0.00	0.00	1.12	0.79	0.39
K02405	RNA polymerase sigma factor for flagellar operon FliA [NA]	0.00	0.00	0.00	1.12	0.27	0.10
K02396	flagellar hook-associated protein 1 FlgK [NA]	0.00	0.00	0.00	1.11	0.00	0.07
K07773	two-component system, OmpR family, aerobic respiration control protein ArcA	0.00	0.00	0.00	1.10	0.29	0.76
K08316	ribosomal RNA small subunit methyltransferase D [EC:2.1.1.52] [EC:2.1.1.52]	0.00	0.00	0.00	1.06	0.66	0.00
K06173	tRNA pseudouridine synthase A [EC:5.4.99.12] [EC:5.4.99.12]	0.00	0.00	0.00	1.04	0.25	0.00
K02414	flagellar hook-length control protein FliK [NA]	0.00	0.00	0.00	1.04	0.32	0.00
K01755	argininosuccinate lyase [EC:4.3.2.1] [EC:4.3.2.1]	0.00	0.17	0.00	1.04	0.97	0.00
K00411	ubiquinol-cytochrome c reductase iron-sulfur subunit [EC:1.10.2.2] [EC:1.10.2	0.00	0.00	0.00	1.04	0.65	0.72
K02407	flagellar hook-associated protein 2 [NA]	0.00	0.00	0.00	1.02	0.14	1.16
K07305	peptide-methionine (R)-S-oxide reductase [EC:1.8.4.12] [EC:1.8.4.12]	0.00	0.00	0.00	1.02	0.00	0.17
K00030	isocitrate dehydrogenase (NAD+) [EC:1.1.1.41] [EC:1.1.1.41]	0.00	0.00	0.00	1.01	0.00	0.07
K07304	peptide-methionine (S)-S-oxide reductase [EC:1.8.4.11]	0.00	0.28	0.00	0.99	0.46	0.08
K00405	cb-type cytochrome c oxidase subunit II [EC:1.9.3.1] [EC:1.9.3.1]	0.00	0.00	0.00	0.99	0.31	0.34
K00931	glutamate 5-kinase [EC:2.7.2.11] [EC:2.7.2.11]	0.00	0.00	0.00	0.91	0.00	0.13
K02301	protoheme IX farnesyltransferase [EC:2.5.1.-] [EC:2.5.1.-]	0.00	0.00	0.00	0.91	0.00	0.00
K02393	flagellar L-ring protein precursor FlgH [NA]	0.00	0.00	0.00	0.91	0.29	0.00
K02388	flagellar basal-body rod protein FlgC [NA]	0.00	0.00	0.00	0.90	1.69	0.00
K01571	oxaloacetate decarboxylase, alpha subunit [EC:4.1.1.3] [EC:4.1.1.3]	0.00	0.00	0.00	0.90	0.21	0.16
K02389	flagellar basal-body rod modification protein FlgD [NA]	0.00	0.00	0.00	0.89	0.56	0.00
K02454	general secretion pathway protein E [NA]	0.00	0.00	0.00	0.89	0.12	0.18
K04088	membrane protease subunit HflK [EC:3.4.-.-] [EC:3.4.-.-]	0.00	0.20	0.00	0.88	0.66	0.24
K02415	flagellar FliL protein [NA]	0.00	0.00	0.00	0.87	0.41	0.15
K03410	chemotaxis protein CheC [NA]	0.00	0.00	0.00	0.87	0.00	0.00
K02395	flagellar protein FlgJ [NA]	0.00	0.00	0.00	0.83	0.39	0.00
K00003	homoserine dehydrogenase [EC:1.1.1.3]	0.00	0.00	0.00	0.83	0.08	0.06
K02409	flagellar M-ring protein FliF [NA]	0.00	0.00	0.00	0.83	0.00	0.00
K03442	small conductance mechanosensitive ion channel, MscS family [NA]	0.00	0.00	0.00	0.83	0.00	0.09
K03071	preprotein translocase SecB subunit [NA]	0.00	0.00	0.00	0.81	0.00	0.14
K00991	2-C-methyl-D-erythritol 4-phosphate cytidylyltransferase [EC:2.7.7.60] [EC:2.	0.00	0.00	0.00	0.80	0.00	0.00
K02110	F-type H+-transporting ATPase subunit c [EC:3.6.3.14] [EC:3.6.3.14]	0.00	0.00	0.00	0.80	0.00	1.66
K02168	high-affinity choline transport protein [NA]	0.00	0.00	0.00	0.80	0.19	0.45
K00641	homoserine O-acetyltransferase [EC:2.3.1.31] [EC:2.3.1.31]	0.00	0.00	0.00	0.78	0.00	0.07
K02416	flagellar motor switch protein FliM [NA]	0.00	0.00	0.00	0.76	0.00	0.26
K02199	cytochrome c biogenesis protein CcmG, thiol:disulfide interchange protein Dsb	0.00	0.00	0.00	0.74	0.35	0.13
K02394	flagellar P-ring protein precursor FlgI [NA]	0.00	0.00	0.00	0.74	0.17	0.00

Data represent the number of sequence hits to each target ortholog per 10,000 reads, normalized to the gene size (in base pairs) of each specific ortholog.

Table S1. Idiomarinaceae specific KEGG orthologues in control and treatment cDNAs

K05589	cell division protein FtsB [NA]	0.00	0.00	0.00	0.73	0.00	0.00
K01501	nitrilase [EC:3.5.5.1] [EC:3.5.5.1]	0.00	0.00	0.00	0.73	0.00	0.00
K01897	long-chain acyl-CoA synthetase [EC:6.2.1.3] [EC:6.2.1.3]	0.00	0.00	0.00	0.73	0.34	0.21
K02275	cytochrome c oxidase subunit II [EC:1.9.3.1] [EC:1.9.3.1]	0.00	0.00	0.00	0.73	0.00	0.00
K00260	glutamate dehydrogenase [EC:1.4.1.2] [EC:1.4.1.2]	0.00	0.00	0.00	0.72	0.17	0.25
K05808	putative sigma-54 modulation protein [NA]	0.00	0.00	0.58	0.72	0.00	0.24
K02258	cytochrome c oxidase subunit XI assembly protein [NA]	0.00	0.00	0.00	0.71	1.01	0.25
K00382	dihydrolipoamide dehydrogenase [EC:1.8.1.4] [EC:1.8.1.4]	0.00	0.00	0.00	0.71	0.27	0.35
K07507	putative Mg2+ transporter-C (MgtC) family protein [NA]	0.00	0.00	0.00	0.71	0.33	0.12
K03570	rod shape-determining protein MreC [NA]	0.00	0.00	0.00	0.71	0.22	0.00
K03089	RNA polymerase sigma-32 factor [NA]	0.00	0.00	0.00	0.71	0.44	0.41
K06178	ribosomal large subunit pseudouridine synthase B [EC:5.4.99.12] [EC:5.4.99.:	0.00	0.00	0.00	0.71	0.89	0.16
K03413	two-component system, chemotaxis family, response regulator CheY [NA]	0.00	0.00	0.00	0.70	0.99	0.00
K02276	cytochrome c oxidase subunit III [EC:1.9.3.1] [EC:1.9.3.1]	0.54	0.00	0.00	0.70	0.22	0.24
K03684	ribonuclease D [EC:3.1.13.5] [EC:3.1.13.5]	0.00	0.00	0.00	0.69	0.00	0.24
K01496	phosphoribosyl-AMP cyclohydrolase [EC:3.5.4.19]	0.00	0.00	0.00	0.67	0.32	0.11
K01920	glutathione synthase [EC:6.3.2.3] [EC:6.3.2.3]	0.00	0.00	0.00	0.65	0.00	0.00
K00627	pyruvate dehydrogenase E2 component (dihydrolipoamide acetyltransferase) [	0.00	0.00	0.00	0.65	0.00	0.80
K00412	ubiquinol-cytochrome c reductase cytochrome b subunit [EC:1.10.2.2] [EC:1.1	0.00	0.00	0.00	0.64	1.36	0.11
K00573	protein-L-isoaspartate(D-aspartate) O-methyltransferase [EC:2.1.1.77] [EC:2.	0.00	0.00	0.00	0.63	0.00	0.00
K02040	phosphate transport system substrate-binding protein [NA]	0.00	0.00	0.00	0.62	0.59	2.74
K00257	putative acyl-CoA dehydrogenase protein (EC:1.3.99.-)	0.00	0.14	0.00	0.62	1.29	0.69
K01914	aspartate--ammonia ligase [EC:6.3.1.1] [EC:6.3.1.1]	0.00	0.00	0.00	0.62	0.39	0.00
K03919	alkylated DNA repair protein [EC:1.14.11.-] [EC:1.14.11.-]	0.00	0.00	0.00	0.62	0.29	0.65
K07684	two-component system, NarL family, nitrate/nitrite response regulator NarL [N	0.00	0.00	0.00	0.62	0.00	0.00
K07685	two-component system, NarL family, nitrate/nitrite response regulator NarP [N	0.00	0.00	0.00	0.62	0.87	0.21
K05560	multicomponent+:H+ antiporter subunit C [NA]	0.00	0.00	0.00	0.60	0.00	0.20
K02488	two-component system, PleD related family, response regulator [NA]	0.00	0.00	0.00	0.60	0.00	0.42
K02982	small subunit ribosomal protein S3 [NA]	0.00	0.00	0.00	0.60	1.69	0.31
K01424	L-asparaginase [EC:3.5.1.1] [EC:3.5.1.1]	0.00	0.00	0.00	0.60	0.00	0.14
K02003	ABC-type transporter, ATP-binding protein	0.00	0.00	0.00	0.59	0.28	0.10
K08363	mercuric ion transport protein [NA]	0.00	0.00	0.00	0.59	0.00	0.00
K03569	rod shape-determining protein MreB and related proteins [NA]	0.00	0.00	0.00	0.59	0.00	0.00
K02410	flagellar motor switch protein FliG [NA]	0.00	0.00	0.00	0.58	0.18	0.00
K01915	glutamine synthetase [EC:6.3.1.2] [EC:6.3.1.2]	0.00	0.00	0.00	0.58	0.14	0.15
K00022	3-hydroxyacyl-CoA dehydrogenase [EC:1.1.1.35]	0.00	0.00	0.00	0.57	0.89	0.62
K02386	flagella basal body P-ring formation protein FlgA [NA]	0.00	0.00	0.00	0.56	0.00	0.00
K00930	acetylglutamateinase [EC:2.7.2.8] [EC:2.7.2.8]	0.00	0.00	0.00	0.56	0.00	0.00
K03072	preprotein translocase SecD subunit [NA]	0.00	0.00	0.00	0.55	0.21	0.11
K00241	succinate dehydrogenase cytochrome b-556 subunit [EC:1.3.99.1] [EC:1.3.99	0.00	0.00	0.00	0.55	0.52	0.38
K01908	propionyl-CoA synthetase [EC:6.2.1.17] [EC:6.2.1.17]	0.00	0.00	0.00	0.54	0.00	0.00
K02411	flagellar assembly protein FliH [NA]	0.00	0.00	0.00	0.54	0.51	0.00
K03414	chemotaxis protein CheZ [NA]	0.00	0.00	0.00	0.54	0.26	0.19
K00507	stearoyl-CoA desaturase (delta-9 desaturase) [EC:1.14.19.1] [EC:1.14.19.1]	0.42	0.00	0.00	0.54	0.34	0.37
K01633	dihydroneopterin aldolase [EC:4.1.2.25] [EC:4.1.2.25]	0.00	0.00	0.00	0.54	0.00	0.00
K00979	3-deoxy-manno-octulosonate cytidylyltransferase (CMP-KDO synthetase) [EC:	0.00	0.00	0.00	0.53	1.26	0.28
K01956	carbamoyl-phosphate synthase small subunit [EC:6.3.5.5] [EC:6.3.5.5]	0.00	0.00	0.00	0.53	0.33	0.06
K00806	undecaprenyl pyrophosphate synthetase [EC:2.5.1.31] [EC:2.5.1.31]	0.00	0.00	0.00	0.53	0.00	0.00
K01895	acetyl-CoA synthetase [EC:6.2.1.1] [EC:6.2.1.1]	0.12	0.00	0.00	0.52	0.20	0.00
K01094	phosphatidylglycerophosphatase [EC:3.1.3.27] [EC:3.1.3.27]	0.00	0.00	0.00	0.52	0.00	0.09
K01903	succinyl-CoA synthetase beta subunit [EC:6.2.1.5] [EC:6.2.1.5]	0.00	0.00	0.00	0.52	0.98	0.12
K07567	TdcF protein [NA]	0.00	0.00	0.00	0.52	0.98	0.00
K03496	chromosome partitioning protein [NA]	0.00	0.00	0.00	0.51	0.00	0.27
K00500	phenylalanine-4-hydroxylase [EC:1.14.16.1] [EC:1.14.16.1]	0.00	0.00	0.00	0.51	0.00	0.00
K02399	flagella synthesis protein FlgN [NA]	0.00	0.00	0.00	0.51	0.00	0.17
K00647	3-oxoacyl-[acyl-carrier-protein] synthase I [EC:2.3.1.41] [acyl-carrier-protein	0.00	0.00	0.00	0.50	0.31	0.52
K06603	flagellar protein FlaG [NA]	0.00	0.00	0.00	0.50	0.47	0.00
K01479	formiminoglutamase [EC:3.5.3.8] [EC:3.5.3.8]	0.00	0.00	0.00	0.50	0.31	0.17
K03574	7,8-dihydro-8-oxoguanine triphosphatase [EC:3.6.1.-] [EC:3.6.1.-]	0.00	0.00	0.00	0.49	0.00	0.00
K01586	diaminopimelate decarboxylase [EC:4.1.1.20] [EC:4.1.1.20]	0.19	0.00	0.00	0.48	0.30	0.11
K03732	ATP-dependent RNA helicase RhlB [EC:3.6.1.-] [EC:3.6.1.-]	0.00	0.00	0.00	0.48	0.00	0.05
K03407	two-component system, chemotaxis family, sensorinase CheA [EC:2.7.13.3] [I	0.00	0.00	0.00	0.47	0.18	0.07
K00290	saccharopine dehydrogenase (NAD+, L-lysine forming) [EC:1.5.1.7] [EC:1.5.1	0.00	0.00	0.00	0.47	0.00	0.16
K02654	leader peptidase (prepilin peptidase) / N-methyltransferase [EC:2.1.1.- 3.4.23	0.00	0.27	0.00	0.47	0.00	0.00
K01451	hippurate hydrolase [EC:3.5.1.32] [EC:3.5.1.32]	0.00	0.00	0.00	0.47	0.15	0.11
K01126	glycerophosphoryl diester phosphodiesterase [EC:3.1.4.46] [EC:3.1.4.46]	0.00	0.00	0.00	0.47	0.00	0.24
K01692	enoyl-CoA hydratase [EC:4.2.1.17] [EC:4.2.1.17]	0.00	0.00	0.00	0.46	0.65	0.16
K06189	magnesium and cobalt transporter [NA]	0.00	0.00	0.00	0.46	0.43	0.16
K07740	regulator of sigma D [NA]	0.00	0.00	0.00	0.46	0.43	0.15
K02160	acetyl-CoA carboxylase biotin carboxyl carrier protein [NA]	0.52	0.00	0.00	0.45	0.42	0.15
K02372	3R-hydroxymyristoyl ACP dehydrase [EC:4.2.1.-] [EC:4.2.1.-]	0.00	0.00	0.00	0.45	0.00	0.00
K02200	cytochrome c-type biogenesis protein CcmH [NA]	0.00	0.00	0.00	0.44	0.42	0.46
K00820	glucosamine--fructose-6-phosphate aminotransferase (isomerizing) [EC:2.6.1.	0.00	0.13	0.00	0.44	0.42	0.00
K03499	trk system potassium uptake protein TrkA [NA]	0.00	0.00	0.00	0.44	0.00	0.26
K00794	riboflavin synthase beta chain [EC:2.5.1.-] [EC:2.5.1.-]	0.00	0.50	0.00	0.44	0.82	0.15
K03409	chemotaxis protein CheX [NA]	0.00	0.00	0.00	0.44	0.00	0.15
K03564	peroxiredoxin Q/BCP [EC:1.11.1.15] [EC:1.11.1.15]	0.00	0.00	0.00	0.44	0.00	0.00
K00026	malate dehydrogenase [EC:1.1.1.37] [EC:1.1.1.37]	0.00	0.00	0.00	0.44	0.41	0.38
K01932	gamma-polyglutamic acid synthetase (EC:6.3.2.-)	0.00	0.00	0.00	0.43	0.00	0.00
K02517	lipid A biosynthesis lauroyl acyltransferase [EC:2.3.1.-] [EC:2.3.1.-]	0.00	0.00	0.00	0.43	0.00	0.00
K01007	pyruvate,water dikinase [EC:2.7.9.2] [EC:2.7.9.2]	0.00	0.00	0.00	0.43	0.00	0.03
K03527	4-hydroxy-3-methylbut-2-enyl diphosphate reductase [EC:1.17.1.2] [EC:1.17.	0.00	0.00	0.00	0.43	0.20	0.07
K00406	cb-type cytochrome c oxidase subunit III [EC:1.9.3.1] [EC:1.9.3.1]	0.00	0.00	0.00	0.43	0.20	0.15
K01962	acetyl-CoA carboxylase carboxyl transferase subunit alpha [EC:6.4.1.2] [EC:6.	0.00	0.00	0.00	0.42	0.20	0.37

Data represent the number of sequence hits to each target ortholog per 10,000 reads, normalized to the gene size (in base pairs) of each specific ortholog.

Table S1. Idiomarinaceae specific KEGG orthologues in control and treatment cDNAs

K06180	ribosomal large subunit pseudouridine synthase D [EC:5.4.99.12] [EC:5.4.99..	0.00	0.00	0.00	0.42	0.20	0.00
K03528	cell division protein ZipA [NA]	0.00	0.00	0.00	0.42	0.39	0.14
K01889	phenylalanyl-tRNA synthetase alpha chain [EC:6.1.1.20] [EC:6.1.1.20]	0.00	0.00	0.00	0.41	0.00	0.00
K00912	tetraacyldisaccharide 4'-kinase [EC:2.7.1.130] [EC:2.7.1.130]	0.00	0.00	0.00	0.41	0.39	0.07
K02259	cytochrome c oxidase subunit XV assembly protein [NA]	0.00	0.00	0.00	0.41	0.00	0.00
K01207	beta-N-acetylhexosaminidase [EC:3.2.1.52] [EC:3.2.1.52]	0.00	0.00	0.00	0.41	0.00	0.00
K01947	biotin-[acetyl-CoA-carboxylase] ligase [EC:6.3.4.15]	0.00	0.00	0.00	0.41	0.19	0.00
K00950	2-amino-4-hydroxy-6-hydroxymethyldihydropteridine pyrophosphokinase [EC:	0.00	0.00	0.00	0.40	0.38	0.00
K01716	3-hydroxydecanoyl-[acyl-carrier-protein] dehydratase [EC:4.2.1.60] [acyl-carr	0.00	0.00	0.00	0.40	0.38	0.00
K00133	aspartate-semialdehyde dehydrogenase [EC:1.2.1.11] [EC:1.2.1.11]	0.00	0.00	0.00	0.40	0.00	0.21
K04047	starvation-inducible DNA-binding protein [NA]	0.00	0.00	0.00	0.40	0.74	0.00
K07462	single-stranded-DNA-specific exonuclease [EC:3.1.-.-] [EC:3.1.-.-]	0.46	0.00	0.00	0.40	0.00	0.00
K00648	3-oxoacyl-[acyl-carrier-protein] synthase III [EC:2.3.1.180] [acyl carrier prote	0.00	0.00	0.00	0.40	0.00	0.41
K02536	UDP-3-O-[3-hydroxymyristoyl] glucosamine N-acyltransferase [EC:2.3.1.-] [3-	0.00	0.00	0.00	0.40	0.00	0.21
K02463	general secretion pathway protein N [NA]	0.00	0.00	0.00	0.40	0.00	0.00
K02864	large subunit ribosomal protein L10 [NA]	0.00	0.00	0.00	0.40	0.37	1.09
K03101	signal peptidase II [EC:3.4.23.36] [EC:3.4.23.36]	0.00	0.00	0.00	0.40	0.37	0.00
K01465	dihydroorotase [EC:3.5.2.3] [EC:3.5.2.3]	0.00	0.00	0.00	0.39	0.00	0.20
K00264	glutamate synthase (NADPH) [EC:1.4.1.13] [EC:1.4.1.13]	0.00	0.00	0.00	0.39	0.92	0.00
K01012	biotin synthetase [EC:2.8.1.6] [EC:2.8.1.6]	0.00	0.00	0.00	0.39	0.00	0.00
K01738	cysteine synthase [EC:2.5.1.47]	0.00	0.00	0.00	0.39	0.55	0.00
K01739	cystathionine gamma-synthase [EC:2.5.1.48] [EC:2.5.1.48]	0.00	0.00	0.00	0.39	0.18	0.27
K07322	regulator of cell morphogenesis and NO signaling [NA]	0.00	0.00	0.00	0.38	0.00	0.00
K07323	putative toluene tolerance protein [NA]	0.00	0.00	0.00	0.38	0.36	0.26
K00457	4-hydroxyphenylpyruvate dioxygenase [EC:1.13.11.27] [EC:1.13.11.27]	0.00	0.00	0.00	0.38	0.00	0.00
K03548	putative permease [NA]	0.00	0.00	0.00	0.38	0.53	0.06
K00831	phosphoserine aminotransferase [EC:2.6.1.52] [EC:2.6.1.52]	0.00	0.00	0.15	0.37	0.18	0.00
K00995	CDP-diacylglycerol--glycerol-3-phosphate 3-phosphatidyltransferase [EC:2.7.8	0.00	0.00	0.00	0.37	0.00	0.00
K02337	DNA polymerase III subunit alpha [EC:2.7.7.7] [EC:2.7.7.7]	0.00	0.00	0.00	0.37	0.06	0.04
K02457	general secretion pathway protein H [NA]	0.00	0.00	0.00	0.37	0.69	0.00
K03270	3-deoxy-D-manno-octulosonate 8-phosphate phosphatase (KDO 8-P phosphat	0.00	0.00	0.00	0.37	0.00	0.13
K03281	chloride channel protein, CIC family [NA]	0.00	0.00	0.00	0.37	0.00	0.00
K02338	DNA polymerase III subunit beta [EC:2.7.7.7] [EC:2.7.7.7]	0.00	0.00	0.00	0.37	0.35	0.00
K03782	catalase/peroxidase [EC:1.11.1.6] [EC:1.11.1.6]	0.00	0.00	0.00	0.37	0.26	0.06
K05559	multicomponent+:H+ antiporter subunit A [NA]	0.00	0.00	0.00	0.36	0.00	0.10
K03310	alanine or glycine:cation symporter, AGCS family [NA]	0.00	0.00	0.00	0.36	0.00	0.13
K03470	ribonuclease HII [EC:3.1.26.4] [EC:3.1.26.4]	0.00	0.00	0.00	0.36	0.00	0.00
K07709	two-component system, NtrC family, sensor histidineinase HydH [EC:2.7.13.3	0.00	0.00	0.00	0.36	0.17	0.00
K01873	valyl-tRNA synthetase [EC:6.1.1.9] [EC:6.1.1.9]	0.00	0.00	0.00	0.36	0.00	0.07
K03386	peroxiredoxin (alkyl hydroperoxide reductase subunit C) [EC:1.11.1.15] [EC:1	0.00	0.00	0.00	0.36	0.00	0.00
K08312	ADP-ribose diphosphatase [EC:3.6.1.-] [EC:3.6.1.-]	0.00	0.00	0.00	0.36	0.00	0.00
K03473	erythronate-4-phosphate dehydrogenase [EC:1.1.1.290] [EC:1.1.1.290]	0.00	0.20	0.00	0.35	0.17	0.37
K03181	chorismate--pyruvate lyase [EC:4.1.3.40] [EC:4.1.3.40]	0.00	0.00	0.00	0.35	0.00	0.12
K02501	glutamine amidotransferase [EC:2.4.2.-] [EC:2.4.2.-]	0.00	0.39	0.00	0.35	0.33	0.00
K02504	protein transport protein HofB [NA]	0.00	0.00	0.00	0.35	0.00	0.00
K00252	glutaryl-CoA dehydrogenase [EC:1.3.99.7] [EC:1.3.99.7]	0.00	0.00	0.00	0.35	0.00	0.00
K03531	cell division protein FtsZ [NA]	0.00	0.00	0.00	0.34	0.48	0.29
K02397	flagellar hook-associated protein 3 FlgL [NA]	0.00	0.00	0.00	0.33	0.00	0.00
K03550	holliday junction DNA helicase RuvA [NA]	0.00	0.00	0.00	0.33	0.00	0.00
K00058	D-3-phosphoglycerate dehydrogenase [EC:1.1.1.95] [EC:1.1.1.95]	0.00	0.00	0.00	0.33	0.16	0.00
K02455	general secretion pathway protein F [NA]	0.00	0.00	0.00	0.33	0.00	0.06
K01662	1-deoxy-D-xylulose-5-phosphate synthase [EC:2.2.1.7] [EC:2.2.1.7]	0.00	0.00	0.00	0.33	0.10	0.08
K03296	hydrophobic/amphiphilic exporter-1 (mainly G- bacteria), HAE1 family [NA]	0.00	0.00	0.00	0.33	0.55	0.16
K01892	histidyl-tRNA synthetase [EC:6.1.1.21] [EC:6.1.1.21]	0.00	0.00	0.00	0.33	0.15	0.23
K02453	general secretion pathway protein D [NA]	0.00	0.00	0.00	0.32	0.20	0.07
K01945	phosphoribosylamine--glycine ligase [EC:6.3.4.13] [EC:6.3.4.13]	0.19	0.00	0.00	0.32	0.45	0.00
K00800	3-phosphoshikimate 1-carboxyvinyltransferase [EC:2.5.1.19]	0.00	0.00	0.00	0.31	0.00	0.00
K01689	enolase [EC:4.2.1.11] [EC:4.2.1.11]	0.00	0.00	0.00	0.31	0.44	0.00
K02198	cytochrome c-type biogenesis protein CcmF [NA]	0.00	0.00	0.00	0.31	0.39	0.14
K01560	2-haloacid dehalogenase [EC:3.8.1.2] [EC:3.8.1.2]	0.00	0.00	0.00	0.31	0.00	0.00
K00013	histidinol dehydrogenase [EC:1.1.1.23] [EC:1.1.1.23]	0.00	0.00	0.00	0.31	0.14	0.00
K01807	ribose 5-phosphate isomerase A [EC:5.3.1.6] [EC:5.3.1.6]	0.00	0.00	0.00	0.30	0.00	0.00
K01783	ribulose-phosphate 3-epimerase [EC:5.1.3.1] [EC:5.1.3.1]	0.00	0.00	0.00	0.30	0.00	0.00
K02412	flagellum-specific ATP synthase [EC:3.6.3.14] [EC:3.6.3.14]	0.00	0.00	0.00	0.30	0.00	0.05
K02194	heme exporter membrane protein CcmB [NA]	0.00	0.00	0.00	0.30	0.00	0.00
K01779	aspartate racemase [EC:5.1.1.13] [EC:5.1.1.13]	0.00	0.00	0.00	0.30	0.28	0.00
K00128	aldehyde dehydrogenase (NAD+) [EC:1.2.1.3] [EC:1.2.1.3]	0.00	0.00	0.00	0.30	0.14	0.00
K03287	outer membrane factor, OMF family [NA]	0.00	0.00	0.00	0.29	0.69	0.10
K10126	two-component system, NtrC family, C4-dicarboxylate transport response regu	0.00	0.00	0.00	0.29	0.00	0.00
K01925	UDP-N-acetylmuramoylalanine--D-glutamate ligase [EC:6.3.2.9] [EC:6.3.2.9]	0.00	0.00	0.00	0.29	0.14	0.00
K00684	leucyl/phenylalanyl-tRNA--protein transferase [EC:2.3.2.6] [EC:2.3.2.6]	0.00	0.00	0.00	0.29	0.00	0.00
K05827	lysine biosynthesis protein LysX [NA]	0.00	0.00	0.00	0.29	0.00	0.00
K10805	acyl-CoA thioesterase II [EC:3.1.2.-] [EC:3.1.2.-]	0.00	0.00	0.00	0.29	0.00	0.40
K02450	general secretion pathway protein A [NA]	0.00	0.00	0.00	0.29	0.14	0.05
K02479	two-component system, NarL family, response regulator [NA]	0.00	0.00	0.00	0.28	0.00	0.00
K02400	flagellar biosynthesis protein FlhA [NA]	0.00	0.00	0.00	0.28	0.27	0.00
K00404	cb-type cytochrome c oxidase subunit I [EC:1.9.3.1] [EC:1.9.3.1]	0.00	0.00	0.00	0.28	0.13	0.00
K02342	DNA polymerase III subunit epsilon [EC:2.7.7.7] [EC:2.7.7.7]	0.00	0.00	0.00	0.28	0.00	0.00
K07665	two-component system, OmpR family, copper resistance phosphate regulon re	0.00	0.00	0.00	0.28	0.00	0.00
K06168	bifunctional enzyme involved in thiolation and methylation of tRNA [NA]	0.00	0.00	0.00	0.28	0.00	0.59
K03474	pyridoxine 5-phosphate synthase [EC:2.6.99.2] [EC:2.6.99.2]	0.00	0.00	0.00	0.28	0.00	0.00
K00568	3-demethylubiquinone-9 3-methyltransferase [EC:2.1.1.64] [EC:2.1.1.64]	0.00	0.00	0.00	0.28	0.26	0.00
K03119	taurine dioxygenase [EC:1.14.11.17] [EC:1.14.11.17]	0.00	0.00	0.00	0.28	0.00	0.00
K01076	palmitoyl-CoA hydrolase (EC:3.1.2.2)	0.00	0.00	0.00	0.28	0.00	0.00

Data represent the number of sequence hits to each target ortholog per 10,000 reads, normalized to the gene size (in base pairs) of each specific ortholog.

Table S1. Idiomarinaceae specific KEGG orthologues in control and treatment cDNAs

K01814	phosphoribosylformimino-5-aminoimidazole carboxamide ribotide isomerase [E	0.00	0.00	0.00	0.28	0.00	0.00
K00029	malate dehydrogenase (oxaloacetate-decarboxylating)(NADP+) [EC:1.1.1.40]	0.00	0.00	0.00	0.28	0.00	0.00
K00655	1-acyl-sn-glycerol-3-phosphate acyltransferase [EC:2.3.1.51] [EC:2.3.1.51]	0.00	0.00	0.00	0.28	0.00	0.19
K01619	deoxyribose-phosphate aldolase [EC:4.1.2.4] [EC:4.1.2.4]	0.00	0.00	0.00	0.27	0.26	0.00
K03787	5'-nucleotidase [EC:3.1.3.5] [EC:3.1.3.5]	0.00	0.00	0.00	0.27	0.51	0.00
K01719	uroporphyrinogen-III synthase [EC:4.2.1.75] [EC:4.2.1.75]	0.00	0.00	0.00	0.27	0.00	0.19
K00532	ferredoxin hydrogenase [EC:1.12.7.2] [EC:1.12.7.2]	0.00	0.00	0.00	0.27	0.00	0.09
K01069	hydroxyacylglutathione hydrolase [EC:3.1.2.6] [EC:3.1.2.6]	0.00	0.00	0.00	0.27	0.00	0.00
K00764	amidophosphoribosyltransferase [EC:2.4.2.14] [EC:2.4.2.14]	0.00	0.00	0.00	0.27	0.00	0.00
K00677	UDP-N-acetylglucosamine acyltransferase [EC:2.3.1.129] [EC:2.3.1.129]	0.00	0.00	0.00	0.27	0.00	0.09
K01834	phosphoglycerate mutase [EC:5.4.2.1] [EC:5.4.2.1]	0.00	0.00	0.00	0.26	0.13	0.05
K00525	ribonucleoside-diphosphate reductase alpha chain [EC:1.17.4.1] [EC:1.17.4.1	0.00	0.00	0.00	0.26	0.00	0.00
K02500	cyclase HisF [EC:4.1.3.-] [EC:4.1.3.-]	0.00	0.00	0.00	0.26	0.00	0.09
K06153	undecaprenyl-diphosphatase [EC:3.6.1.27] [EC:3.6.1.27]	0.00	0.00	0.00	0.26	0.72	0.09
K00606	3-methyl-2-oxobutanoate hydroxymethyltransferase [EC:2.1.2.11] [EC:2.1.2.	0.00	0.00	0.00	0.26	0.00	0.00
K02065	putative ABC transport system ATP-binding protein [NA]	0.00	0.00	0.00	0.26	0.00	0.09
K00796	dihydropteroate synthase [EC:2.5.1.15] [EC:2.5.1.15]	0.00	0.00	0.00	0.25	0.00	0.00
K01525	bis(5'-nucleosyl)-tetraphosphatase (symmetrical) [EC:3.6.1.41] [EC:3.6.1.41]	0.00	0.00	0.00	0.25	0.00	0.08
K00286	pyrroline-5-carboxylate reductase [EC:1.5.1.2] [EC:1.5.1.2]	0.00	0.00	0.00	0.25	0.23	0.00
K00674	2,3,4,5-tetrahydropyridine-2-carboxylate N-succinyltransferase [EC:2.3.1.117	0.29	0.00	0.00	0.25	0.00	0.26
K01835	phosphoglucomutase [EC:5.4.2.2] [EC:5.4.2.2]	0.00	0.00	0.00	0.25	0.35	0.13
K00640	serine O-acetyltransferase [EC:2.3.1.30] [EC:2.3.1.30]	0.00	0.00	0.00	0.25	0.23	0.17
K00998	phosphatidylserine synthase [EC:2.7.8.8] [EC:2.7.8.8]	0.00	0.28	0.00	0.25	0.00	0.25
K01580	glutamate decarboxylase [EC:4.1.1.15] [EC:4.1.1.15]	0.00	0.00	0.00	0.25	0.12	0.13
K01652	acetolactate synthase I/II/III large subunit [EC:2.2.1.6] [EC:2.2.1.6]	0.00	0.00	0.10	0.24	0.12	0.09
K00014	shikimate 5-dehydrogenase [EC:1.1.1.25]	0.00	0.00	0.00	0.24	0.00	0.00
K01627	2-dehydro-3-deoxyphosphooctonate aldolase (KDO 8-P synthase) [EC:2.5.1.5!	0.00	0.00	0.00	0.24	0.00	0.00
K01886	glutaminyl-tRNA synthetase [EC:6.1.1.18] [EC:6.1.1.18]	0.00	0.00	0.00	0.24	0.23	0.04
K02341	DNA polymerase III subunit delta' [EC:2.7.7.7] [EC:2.7.7.7]	0.00	0.00	0.00	0.24	0.22	0.08
K09686	antibiotic transport system permease protein [NA]	0.00	0.00	0.00	0.23	0.00	0.00
K01966	propionyl-CoA carboxylase beta chain [EC:6.4.1.3] [EC:6.4.1.3]	0.00	0.00	0.19	0.23	0.00	0.00
K04087	membrane protease subunit HflC [EC:3.4.-.-] [EC:3.4.-.-]	0.00	0.00	0.00	0.23	1.28	0.55
K05844	ribosomal protein S6 modification protein [NA]	0.00	0.00	0.00	0.23	0.00	0.00
K01081	5'-nucleotidase [EC:3.1.3.5] [EC:3.1.3.5]	0.00	0.00	0.00	0.22	0.21	0.08
K01480	agmatinase [EC:3.5.3.11] [EC:3.5.3.11]	0.00	0.00	0.00	0.22	0.21	0.08
K01921	D-alanine-D-alanine ligase [EC:6.3.2.4] [EC:6.3.2.4]	0.00	0.00	0.00	0.22	0.00	0.00
K03644	lipoic acid synthetase [EC:2.8.1.8] [EC:2.8.1.8]	0.00	0.00	0.00	0.22	0.00	0.07
K02557	chemotaxis protein MotB	0.00	0.00	0.00	0.22	0.00	0.00
K00164	2-oxoglutarate dehydrogenase E1 component [EC:1.2.4.2] [EC:1.2.4.2]	0.00	0.00	0.00	0.22	0.07	0.18
K09699	2-oxoisovalerate dehydrogenase E2 component (dihydrolipoyl transacylase) [E	0.00	0.00	0.00	0.22	0.00	0.30
K01870	isoleucyl-tRNA synthetase [EC:6.1.1.5] [EC:6.1.1.5]	0.00	0.00	0.00	0.22	0.34	0.10
K01425	glutaminase [EC:3.5.1.2] [EC:3.5.1.2]	0.00	0.00	0.00	0.21	0.00	0.00
K01697	cystathionine beta-synthase [EC:4.2.1.22] [EC:4.2.1.22]	0.00	0.00	0.00	0.21	0.20	0.00
K03087	RNA polymerase nonessential primary-like sigma factor [NA]	0.00	0.48	0.00	0.21	0.00	0.07
K02460	general secretion pathway protein [NA]	0.00	0.00	0.00	0.21	0.00	0.00
K01238	membrane-bound lytic murein transglycosylase B (EC:3.2.1.-)	0.00	0.00	0.00	0.21	0.19	0.14
K00946	thiamine-monophosphateinase [EC:2.7.4.16] [EC:2.7.4.16]	0.00	0.00	0.00	0.21	0.00	0.00
K00057	glycerol-3-phosphate dehydrogenase (NAD(P)+) [EC:1.1.1.94] [EC:1.1.1.94]	0.00	0.00	0.00	0.20	0.00	0.21
K01551	arsenite-transporting ATPase [EC:3.6.3.16] [EC:3.6.3.16]	0.00	0.00	0.00	0.20	0.00	0.14
K01972	DNA ligase (NAD+) [EC:6.5.1.2] [EC:6.5.1.2]	0.00	0.00	0.00	0.20	0.00	0.03
K01067	acetyl-CoA hydrolase [EC:3.1.2.1] [EC:3.1.2.1]	0.00	0.00	0.00	0.20	0.19	0.00
K02343	DNA polymerase III subunit gamma/tau [EC:2.7.7.7] [EC:2.7.7.7]	0.00	0.00	0.00	0.20	0.19	0.03
K02669	twitching motility protein PilT [NA]	0.00	0.00	0.00	0.20	0.00	0.00
K01041	4-hydroxybutyrate coenzyme A transferase (EC:2.8.3.-)	0.00	0.00	0.00	0.20	0.00	0.00
K00609	aspartate carbamoyltransferase catalytic subunit [EC:2.1.3.2] [EC:2.1.3.2]	0.00	0.00	0.00	0.20	0.37	0.34
K06176	tRNA pseudouridine synthase D [EC:5.4.99.12] [EC:5.4.99.12]	0.00	0.00	0.00	0.19	0.00	0.00
K01933	phosphoribosylformylglycinamidine cyclo-ligase [EC:6.3.3.1] [EC:6.3.3.1]	0.00	0.00	0.00	0.19	0.36	0.00
K07568	S-adenosylmethionine:tRNA ribosyltransferase-isomerase [EC:5.-.-.-] [EC:5.-.	0.00	0.00	0.00	0.19	0.00	0.00
K07239	heavy-metal exporter, HME family [NA]	0.00	0.00	0.00	0.19	0.24	0.00
K00294	1-pyrroline-5-carboxylate dehydrogenase [EC:1.5.1.12]	0.00	0.00	0.00	0.19	0.06	0.31
K01624	fructose-bisphosphate aldolase, class II [EC:4.1.2.13] [EC:4.1.2.13]	0.00	0.00	0.00	0.19	0.18	0.00
K00025	malate dehydrogenase [EC:1.1.1.37] [EC:1.1.1.37]	0.00	0.11	0.00	0.19	0.18	0.00
K01000	phospho-N-acetylmuramoyl-pentapeptide-transferase [EC:2.7.8.13] [EC:2.7.8	0.00	0.00	0.00	0.19	0.00	0.06
K00817	histidinol-phosphate aminotransferase [EC:2.6.1.9] [EC:2.6.1.9]	0.00	0.00	0.00	0.19	0.00	0.00
K01736	chorismate synthase [EC:4.2.3.5] [EC:4.2.3.5]	0.00	0.00	0.00	0.19	0.00	0.19
K01775	alanine racemase [EC:5.1.1.1] [EC:5.1.1.1]	0.00	0.00	0.00	0.19	0.18	0.06
K00951	GTP pyrophosphokinase [EC:2.7.6.5] [EC:2.7.6.5]	0.00	0.00	0.00	0.19	0.17	0.03
K03526	4-hydroxy-3-methylbut-2-en-1-yl diphosphate synthase [EC:1.17.4.3] [EC:1.1	0.00	0.00	0.00	0.18	0.17	0.06
K05837	rod shape determining protein RodA [NA]	0.00	0.00	0.00	0.18	0.00	0.00
K00082	5-amino-6-(5-phosphoribosylamino)uracil reductase [EC:1.1.1.193]	0.00	0.00	0.00	0.18	0.17	0.00
K01533	Cu2+-exporting ATPase [EC:3.6.3.4] [EC:3.6.3.4]	0.00	0.00	0.07	0.18	0.09	0.00
K08484	phosphotransferase system, enzyme I, PtsP [EC:2.7.3.9] [EC:2.7.3.9]	0.10	0.00	0.00	0.18	0.67	0.12
K04487	cysteine desulfurase [EC:2.8.1.7] [EC:2.8.1.7]	0.00	0.00	0.00	0.18	0.66	0.06
K03412	protein-glutamate methylesterase, two-component system, chemotaxis family	0.00	0.00	0.00	0.18	0.66	0.18
K00012	UDPglucose 6-dehydrogenase [EC:1.1.1.22] [EC:1.1.1.22]	0.00	0.00	0.00	0.18	0.00	0.06
K03185	2-octaprenyl-6-methoxyphenol hydroxylase [EC:1.14.13.-] [EC:1.14.13.-]	0.41	0.00	0.00	0.18	0.16	0.00
K00927	phosphoglycerateinase [EC:2.7.2.3] [EC:2.7.2.3]	0.00	0.00	0.00	0.17	0.32	0.00
K01843	lysine 2,3-aminomutase [EC:5.4.3.2] [EC:5.4.3.2]	0.00	0.00	0.00	0.17	0.16	0.00
K01866	tyrosyl-tRNA synthetase [EC:6.1.1.1] [EC:6.1.1.1]	0.00	0.00	0.00	0.17	0.00	0.00
K03317	concentrative nucleoside transporter, CNT family [NA]	0.00	0.00	0.00	0.17	0.16	0.00
K03320	ammonium transporter, Amt family [NA]	0.00	0.00	0.00	0.17	0.16	0.06
K01940	argininosuccinate synthase [EC:6.3.4.5] [EC:6.3.4.5]	0.39	0.00	0.00	0.17	0.63	0.00
K03466	DNA segregation ATPase FtsK/SpoIIIE, S-DNA-T family [NA]	0.00	0.00	0.00	0.17	0.08	0.03
K01468	imidazolonepropionase [EC:3.5.2.7] [EC:3.5.2.7]	0.00	0.00	0.00	0.17	0.00	0.00

Data represent the number of sequence hits to each target ortholog per 10,000 reads, normalized to the gene size (in base pairs) of each specific ortholog.

Table S1. Idiomarinaceae specific KEGG orthologues in control and treatment cDNAs

K03588	cell division protein FtsW [NA]	0.00	0.00	0.00	0.17	0.16	0.06
K00600	glycine hydroxymethyltransferase [EC:2.1.2.1] [EC:2.1.2.1]	0.00	0.00	0.00	0.16	0.30	0.00
K03628	transcription termination factor Rho [NA]	0.00	0.00	0.00	0.16	0.15	0.39
K02492	glutamyl-tRNA reductase [EC:1.2.1.70] [EC:1.2.1.70]	0.00	0.00	0.00	0.16	0.00	0.11
K00631	glycerol-3-phosphate O-acyltransferase [EC:2.3.1.15] [EC:2.3.1.15]	0.00	0.00	0.00	0.16	0.15	0.08
K01875	seryl-tRNA synthetase [EC:6.1.1.11] [EC:6.1.1.11]	0.00	0.00	0.00	0.16	0.00	0.43
K01927	dihydrofolate synthase [EC:6.3.2.12]	0.00	0.00	0.00	0.16	0.00	0.27
K03885	NADH dehydrogenase [EC:1.6.99.3] [EC:1.6.99.3]	0.00	0.00	0.00	0.16	0.00	0.00
K07636	two-component system, OmpR family, phosphate regulon sensor histidineinas	0.00	0.00	0.00	0.16	1.03	0.11
K01869	leucyl-tRNA synthetase [EC:6.1.1.4] [EC:6.1.1.4]	0.00	0.00	0.00	0.16	0.07	0.03
K01077	alkaline phosphatase [EC:3.1.3.1] [EC:3.1.3.1]	0.00	0.00	0.00	0.16	0.15	0.11
K01872	alanyl-tRNA synthetase [EC:6.1.1.7] [EC:6.1.1.7]	0.18	0.00	0.00	0.16	0.22	0.03
K07638	two-component system, OmpR family, osmolarity sensor histidineinase EnvZ [	0.00	0.00	0.00	0.16	0.29	0.05
K01129	dGTPase [EC:3.1.5.1] [EC:3.1.5.1]	0.00	0.00	0.00	0.15	0.00	0.00
K03500	ribosomal RNA small subunit methyltransferase B [EC:2.1.1.-] [EC:2.1.1.-]	0.00	0.00	0.00	0.15	0.00	0.37
K03498	trk system potassium uptake protein TrkH [NA]	0.00	0.00	0.00	0.15	0.14	0.00
K09760	DNA recombination protein RmuC [NA]	0.00	0.00	0.00	0.15	0.00	0.16
K00163	pyruvate dehydrogenase E1 component [EC:1.2.4.1] [EC:1.2.4.1]	0.00	0.00	0.00	0.15	0.00	0.18
K01486	adenine deaminase [EC:3.5.4.2] [EC:3.5.4.2]	0.00	0.00	0.00	0.15	0.00	0.00
K01488	adenosine deaminase [EC:3.5.4.4] [EC:3.5.4.4]	0.00	0.00	0.00	0.15	0.00	0.21
K01946	biotin carboxylase [EC:6.3.4.14]	0.00	0.00	0.00	0.15	0.00	0.00
K00383	glutathione reductase (NADPH) [EC:1.8.1.7] [EC:1.8.1.7]	0.00	0.00	0.00	0.15	0.14	0.10
K01626	3-deoxy-7-phosphoheptulonate synthase [EC:2.5.1.54] [EC:2.5.1.54]	0.00	0.00	0.00	0.15	0.00	0.26
K00161	pyruvate dehydrogenase E1 component subunit alpha [EC:1.2.4.1] [EC:1.2.4.	0.00	0.00	0.00	0.15	0.00	0.00
K03294	basic amino acid/polyamine antiporter, APA family [NA]	0.00	0.00	0.00	0.15	0.00	0.31
K01893	asparaginyl-tRNA synthetase [EC:6.1.1.22] [EC:6.1.1.22]	0.00	0.00	0.00	0.15	0.00	0.10
K07645	two-component system, OmpR family, sensor histidineinase QseC [EC:2.7.13.:	0.00	0.00	0.00	0.14	0.00	0.05
K07648	two-component system, OmpR family, aerobic respiration control sensor histid	0.00	0.00	0.00	0.14	0.00	0.00
K00982	glutamate-ammonia-ligase adenylyltransferase [EC:2.7.7.42] [EC:2.7.7.42]	0.00	0.00	0.00	0.14	0.07	0.17
K00088	IMP dehydrogenase [EC:1.1.1.205] [EC:1.1.1.205]	0.00	0.00	0.00	0.14	0.13	0.00
K06447	succinylglutamic semialdehyde dehydrogenase [EC:1.2.1.71] [EC:1.2.1.71]	0.00	0.00	0.00	0.14	0.13	0.10
K08301	ribonuclease G [EC:3.1.4.-] [EC:3.1.4.-]	0.16	0.00	0.00	0.14	0.00	0.00
K08300	ribonuclease E [EC:3.1.4.-] [EC:3.1.4.-]	0.00	0.00	0.00	0.14	0.07	0.05
K02600	N utilization substance protein A [NA]	0.00	0.00	0.00	0.14	0.64	0.28
K05561	multicomponent+:H+ antiporter subunit D [NA]	0.00	0.00	0.00	0.13	0.00	0.00
K01676	fumarate hydratase, class I [EC:4.2.1.2] [EC:4.2.1.2]	0.00	0.00	0.00	0.13	0.13	0.00
K00658	2-oxoglutarate dehydrogenase E2 component (dihydrolipoamide succinyltransf	0.00	0.00	0.00	0.13	0.37	0.32
K03980	virulence factor [NA]	0.00	0.00	0.00	0.13	0.00	0.00
K03776	aerotaxis receptor [NA]	0.00	0.00	0.00	0.13	0.24	0.09
K07787	Cu(I)/Ag(I) efflux system membrane protein CusA [NA]	0.00	0.00	0.00	0.13	0.00	0.00
K01951	GMP synthase (glutamine-hydrolysing) [EC:6.3.5.2] [EC:6.3.5.2]	0.00	0.00	0.00	0.13	0.24	0.00
K01657	anthranilate synthase component I [EC:4.1.3.27] [EC:4.1.3.27]	0.00	0.00	0.00	0.13	0.00	0.22
K01919	glutamate--cysteine ligase [EC:6.3.2.2] [EC:6.3.2.2]	0.00	0.00	0.00	0.13	0.12	0.04
K02038	phosphate transport system permease protein [NA]	0.00	0.00	0.00	0.12	0.46	0.04
K00166	2-oxoisovalerate dehydrogenase E1 component, alpha subunit [EC:1.2.4.4] [E	0.00	0.00	0.00	0.12	0.45	0.17
K00681	gamma-glutamyltranspeptidase [EC:2.3.2.2] [EC:2.3.2.2]	0.00	0.00	0.00	0.12	0.22	0.20
K03316	monovalent cation:H+ antiporter, CPA1 family [NA]	0.00	0.00	0.00	0.12	0.11	0.12
K03587	cell division protein FtsI (penicillin binding protein 3) [EC:2.4.1.129] [EC:2.4.1	0.00	0.00	0.00	0.12	0.00	0.00
K02316	DNA primase [EC:2.7.7.-] [EC:2.7.7.-]	0.00	0.00	0.00	0.12	0.00	0.36
K00239	succinate dehydrogenase flavoprotein subunit [EC:1.3.99.1] [EC:1.3.99.1]	0.00	0.00	0.00	0.12	0.00	0.20
K03086	RNA polymerase primary sigma factor [NA]	0.00	0.00	0.00	0.11	0.00	0.15
K03703	excinuclease ABC subunit C [NA]	0.00	0.00	0.00	0.11	0.10	0.00
K03654	ATP-dependent DNA helicase RecQ [EC:3.6.1.-] [EC:3.6.1.-]	0.00	0.00	0.00	0.11	0.10	0.00
K03582	exodeoxyribonuclease V beta subunit [EC:3.1.11.5] [EC:3.1.11.5]	0.00	0.00	0.00	0.11	0.20	0.00
K02004	hypothetical protein	0.13	0.00	0.00	0.11	0.30	0.22
K01585	arginine decarboxylase [EC:4.1.1.19] [EC:4.1.1.19]	0.00	0.00	0.00	0.11	0.10	0.07
K04079	molecular chaperone HtpG [NA]	0.00	0.00	0.00	0.11	0.00	0.04
K01868	threonyl-tRNA synthetase [EC:6.1.1.3] [EC:6.1.1.3]	0.00	0.00	0.00	0.11	0.40	0.29
K03798	cell division protease FtsH [EC:3.4.24.-] [EC:3.4.24.-]	0.00	0.00	0.00	0.10	0.20	0.07
K03455	K+ transport system, membrane component	0.00	0.00	0.00	0.10	0.00	0.04
K03578	ATP-dependent helicase HrpA [EC:3.6.1.-] [EC:3.6.1.-]	0.00	0.00	0.00	0.10	0.05	0.00
K00619	amino-acid N-acetyltransferase [EC:2.3.1.1]	0.00	0.00	0.00	0.10	0.10	0.00
K01953	asparagine synthase (glutamine-hydrolysing) [EC:6.3.5.4] [EC:6.3.5.4]	0.00	0.00	0.00	0.10	0.19	0.35
K01874	methionyl-tRNA synthetase [EC:6.1.1.10] [EC:6.1.1.10]	0.00	0.00	0.00	0.10	0.19	0.03
K01879	glycyl-tRNA synthetase beta chain [EC:6.1.1.14] [EC:6.1.1.14]	0.00	0.00	0.00	0.10	0.18	0.07
K03655	ATP-dependent DNA helicase RecG [EC:3.6.1.-] [EC:3.6.1.-]	0.00	0.00	0.00	0.10	0.09	0.14
K03046	DNA-directed RNA polymerase subunit beta' [EC:2.7.7.6] [EC:2.7.7.6]	0.00	0.00	0.04	0.10	0.50	0.15
K00962	polyribonucleotide nucleotidyltransferase [EC:2.7.7.8] [EC:2.7.7.8]	0.00	0.00	0.00	0.09	0.09	0.07
K02401	flagellar biosynthetic protein FlhB [NA]	0.00	0.00	0.00	0.09	0.00	0.00
K05365	penicillin binding protein 1B [EC:2.4.1.129 3.4.-.-] [EC:2.4.1.129 3.4.-.-]	0.00	0.00	0.00	0.09	0.25	0.00
K00117	quinoprotein glucose dehydrogenase [EC:1.1.5.2] [EC:1.1.5.2]	0.00	0.00	0.00	0.09	0.00	0.03
K01529	RecG-like helicase	0.00	0.00	0.00	0.09	0.00	0.06
K03579	ATP-dependent helicase HrpB [EC:3.6.1.-] [EC:3.6.1.-]	0.00	0.00	0.00	0.08	0.00	0.00
K01259	proline iminopeptidase [EC:3.4.11.5] [EC:3.4.11.5]	0.00	0.00	0.00	0.08	0.00	0.00
K00990	[protein-PII] uridylyltransferase [EC:2.7.7.59] [protein-PII] uridylyltransferase	0.00	0.00	0.00	0.08	0.00	0.00
K03580	ATP-dependent helicase HepA [EC:3.6.1.-] [EC:3.6.1.-]	0.00	0.00	0.00	0.08	0.36	0.05
K02335	DNA polymerase I [EC:2.7.7.7] [EC:2.7.7.7]	0.00	0.00	0.00	0.07	0.00	0.02
K01955	carbamoyl-phosphate synthase large subunit [EC:6.3.5.5] [EC:6.3.5.5]	0.00	0.00	0.00	0.06	0.06	0.00
K03529	chromosome segregation protein [NA]	0.07	0.00	0.00	0.06	0.22	0.04
K03583	exodeoxyribonuclease V gamma subunit [EC:3.1.11.5] [EC:3.1.11.5]	0.00	0.00	0.00	0.06	0.06	0.02
K00001	alcohol dehydrogenase [EC:1.1.1.1] [EC:1.1.1.1]	0.00	0.00	0.00	0.00	0.00	0.06
K00020	3-hydroxyisobutyrate dehydrogenase [EC:1.1.1.31] [EC:1.1.1.31]	0.00	0.53	0.00	0.00	0.22	0.08
K00031	isocitrate dehydrogenase [EC:1.1.1.42] [EC:1.1.1.42]	0.00	0.00	0.00	0.00	0.17	0.09
K00059	3-oxoacyl-[acyl-carrier protein] reductase [EC:1.1.1.100] [acyl-carrier protein	0.00	0.00	0.00	0.00	0.00	1.42

Data represent the number of sequence hits to each target ortholog per 10,000 reads, normalized to the gene size (in base pairs) of each specific ortholog.

Table S1. Idiomarinaceae specific KEGG orthologues in control and treatment cDNAs

K00060	threonine 3-dehydrogenase [EC:1.1.1.103] [EC:1.1.1.103]	0.00	0.00	0.00	0.00	0.00	0.14
K00075	UDP-N-acetylmuramate dehydrogenase [EC:1.1.1.158] [EC:1.1.1.158]	0.00	0.00	0.00	0.00	0.00	0.07
K00099	1-deoxy-D-xylulose-5-phosphate reductoisomerase [EC:1.1.1.267] [EC:1.1.1.	0.00	0.00	0.00	0.00	0.32	0.24
K00100	glucose dehydrogenase 1, type B	0.00	0.00	0.00	0.00	0.00	0.12
K00134	glyceraldehyde 3-phosphate dehydrogenase [EC:1.2.1.12] [EC:1.2.1.12]	0.00	0.00	0.00	0.00	0.38	0.21
K00140	methylmalonate-semialdehyde dehydrogenase [EC:1.2.1.27] [EC:1.2.1.27]	0.00	0.00	0.00	0.00	0.00	0.19
K00145	N-acetyl-gamma-glutamyl-phosphate reductase [EC:1.2.1.38] [EC:1.2.1.38]	0.00	0.00	0.00	0.00	0.00	0.14
K00147	glutamate-5-semialdehyde dehydrogenase [EC:1.2.1.41] [EC:1.2.1.41]	0.00	0.00	0.00	0.00	0.61	0.00
K00167	2-oxoisovalerate dehydrogenase E1 component, beta subunit [EC:1.2.4.4] [EC	0.00	0.00	0.00	0.00	0.20	0.00
K00226	dihydroorotate oxidase [EC:1.3.3.1] [EC:1.3.3.1]	0.00	0.00	0.00	0.00	0.19	0.07
K00228	coproporphyrinogen III oxidase [EC:1.3.3.3] [EC:1.3.3.3]	0.00	0.00	0.00	0.00	0.21	0.08
K00249	acyl-CoA dehydrogenase [EC:1.3.99.3] [EC:1.3.99.3]	0.00	0.00	0.00	0.00	0.56	0.20
K00253	isovaleryl-CoA dehydrogenase [EC:1.3.99.10] [EC:1.3.99.10]	0.00	0.00	0.00	0.00	0.16	0.00
K00259	alanine dehydrogenase [EC:1.4.1.1] [EC:1.4.1.1]	0.00	0.00	0.00	0.00	0.00	0.06
K00262	glutamate dehydrogenase (NADP+) [EC:1.4.1.4] [EC:1.4.1.4]	0.00	0.00	0.12	0.00	0.14	0.00
K00278	L-aspartate oxidase [EC:1.4.3.16] [EC:1.4.3.16]	0.00	0.00	0.00	0.00	0.00	0.04
K00282	glycine dehydrogenase subunit 1 [EC:1.4.4.2]	0.00	0.00	0.00	0.00	0.20	0.00
K00288	methylenetetrahydrofolate dehydrogenase (NADP+) [EC:1.5.1.5]	0.00	0.00	0.00	0.00	0.22	0.00
K00322	NAD(P) transhydrogenase [EC:1.6.1.1] [EC:1.6.1.1]	0.00	0.00	0.00	0.00	0.00	0.10
K00341	NADH dehydrogenase I subunit L [EC:1.6.5.3] [EC:1.6.5.3]	0.00	0.00	0.00	0.00	0.00	0.05
K00384	thioredoxin reductase (NADPH) [EC:1.8.1.9] [EC:1.8.1.9]	0.00	0.00	0.00	0.00	0.20	0.15
K00407	cb-type cytochrome c oxidase subunit IV [EC:1.9.3.1] [EC:1.9.3.1]	0.00	0.00	0.00	0.00	0.00	0.38
K00426	cytochrome bd-I oxidase subunit II [EC:1.10.3.-] [EC:1.10.3.-]	0.00	0.47	0.00	0.00	0.57	0.00
K00554	tRNA (guanine-N1-)-methyltransferase [EC:2.1.1.31] [EC:2.1.1.31]	0.00	0.00	0.00	0.00	1.26	1.12
K00556	tRNA (guanosine-2'-O-)-methyltransferase [EC:2.1.1.34] [EC:2.1.1.34]	0.00	0.00	0.00	0.00	0.30	0.22
K00560	thymidylate synthase [EC:2.1.1.45] [EC:2.1.1.45]	0.00	0.00	0.19	0.00	0.00	0.00
K00575	chemotaxis protein methyltransferase CheR [EC:2.1.1.80] [EC:2.1.1.80]	0.28	0.00	0.00	0.00	0.46	0.51
K00599	Unclassified	0.00	0.00	0.00	0.00	0.33	0.00
K00602	phosphoribosylaminoimidazolecarboxamide formyltransferase [EC:2.1.2.3]	0.31	0.00	0.00	0.00	0.00	0.04
K00604	methionyl-tRNA formyltransferase [EC:2.1.2.9] [EC:2.1.2.9]	0.00	0.00	0.00	0.00	0.00	0.07
K00605	aminomethyltransferase [EC:2.1.2.10] [EC:2.1.2.10]	0.00	0.00	0.00	0.00	0.00	0.06
K00611	ornithine carbamoyltransferase [EC:2.1.3.3] [EC:2.1.3.3]	0.00	0.00	0.00	0.00	0.40	0.00
K00615	transketolase [EC:2.2.1.1] [EC:2.2.1.1]	0.00	0.00	0.00	0.00	0.00	0.03
K00626	acetyl-CoA C-acetyltransferase [EC:2.3.1.9] [EC:2.3.1.9]	0.00	0.00	0.00	0.00	0.16	0.00
K00645	[acyl-carrier-protein] S-malonyltransferase [EC:2.3.1.39] [acyl-carrier-protein	0.00	0.00	0.00	0.00	0.00	0.08
K00666	fatty-acyl-CoA synthase [EC:6.2.1.-] [EC:6.2.1.-]	0.00	0.00	0.00	0.00	0.00	0.04
K00673	arginine N-succinyltransferase [EC:2.3.1.109] [EC:2.3.1.109]	0.00	0.00	0.00	0.00	0.75	0.21
K00680	acetyltransferase (EC:2.3.1.-)	0.00	0.00	0.00	0.00	0.00	0.11
K00748	lipid-A-disaccharide synthase [EC:2.4.1.182] [EC:2.4.1.182]	0.00	0.00	0.00	0.00	0.00	0.25
K00758	thymidine phosphorylase [EC:2.4.2.4] [EC:2.4.2.4]	0.00	0.00	0.00	0.00	0.14	0.05
K00759	adenine phosphoribosyltransferase [EC:2.4.2.7] [EC:2.4.2.7]	0.00	0.00	0.00	0.00	0.36	0.40
K00766	anthranilate phosphoribosyltransferase [EC:2.4.2.18] [EC:2.4.2.18]	0.00	0.00	0.00	0.00	0.00	0.07
K00769	xanthine phosphoribosyltransferase [EC:2.4.2.22] [EC:2.4.2.22]	0.00	0.00	0.00	0.00	0.40	0.00
K00789	S-adenosylmethionine synthetase [EC:2.5.1.6] [EC:2.5.1.6]	0.00	0.00	0.00	0.00	0.00	0.18
K00790	UDP-N-acetylglucosamine 1-carboxyvinyltransferase [EC:2.5.1.7] [EC:2.5.1.7]	0.00	0.00	0.00	0.00	0.15	0.11
K00793	riboflavin synthase alpha chain [EC:2.5.1.9] [EC:2.5.1.9]	0.00	0.00	0.00	0.00	0.00	0.23
K00795	geranyltranstransferase [EC:2.5.1.10] [EC:2.5.1.10]	0.00	0.00	0.00	0.00	0.00	0.16
K00812	aspartate aminotransferase [EC:2.6.1.1] [EC:2.6.1.1]	0.00	0.00	0.00	0.00	0.16	0.00
K00821	acetylornithine/N-succinyldiaminopimelate aminotransferase [EC:2.6.1.11 2.6.	0.00	0.00	0.00	0.00	0.83	0.08
K00826	branched-chain amino acid aminotransferase [EC:2.6.1.42] [EC:2.6.1.42]	0.00	0.00	0.00	0.00	0.19	0.00
K00833	adenosylmethionine-8-amino-7-oxononanoate aminotransferase [EC:2.6.1.62]	0.00	0.00	0.00	0.00	0.29	0.05
K00847	fructokinase [EC:2.7.1.4] [EC:2.7.1.4]	0.00	0.00	0.00	0.00	0.15	0.05
K00850	6-phosphofructokinase [EC:2.7.1.11] [EC:2.7.1.11]	0.00	0.00	0.00	0.00	0.19	0.14
K00852	ribokinase [EC:2.7.1.15] [EC:2.7.1.15]	0.00	0.00	0.00	0.00	0.37	0.00
K00855	phosphoribulokinase [EC:2.7.1.19] [EC:2.7.1.19]	0.00	0.00	0.00	0.00	0.21	0.00
K00857	thymidineinase [EC:2.7.1.21] [EC:2.7.1.21]	0.00	0.00	0.00	0.00	0.33	0.00
K00858	NAD+inase [EC:2.7.1.23] [EC:2.7.1.23]	0.00	0.00	0.00	0.00	0.22	0.08
K00861	riboflavininase [EC:2.7.1.26]	0.00	0.00	0.00	0.00	0.21	0.07
K00873	pyruvateinase [EC:2.7.1.40] [EC:2.7.1.40]	0.00	0.00	0.00	0.00	0.00	0.05
K00876	uridineinase [EC:2.7.1.48] [EC:2.7.1.48]	0.00	0.00	0.00	0.00	0.00	0.26
K00891	shikimateinase [EC:2.7.1.71] [EC:2.7.1.71]	0.00	0.00	0.31	0.00	0.00	0.13
K00924	Unclassified	0.00	0.00	0.19	0.00	0.22	0.00
K00939	adenylateinase [EC:2.7.4.3] [EC:2.7.4.3]	0.00	0.00	0.00	0.00	0.00	0.22
K00942	guanylateinase [EC:2.7.4.8] [EC:2.7.4.8]	0.00	0.00	0.00	0.00	0.30	0.00
K00963	UTP--glucose-1-phosphate uridylyltransferase [EC:2.7.7.9] [EC:2.7.7.9]	0.00	0.00	0.00	0.00	0.22	0.00
K00966	mannose-1-phosphate guanylyltransferase [EC:2.7.7.13] [EC:2.7.7.13]	0.00	0.00	0.00	0.00	0.22	0.00
K00981	phosphatidate cytidylyltransferase [EC:2.7.7.41] [EC:2.7.7.41]	0.00	0.00	0.00	0.00	0.44	0.08
K00983	N-acylneuraminate cytidylyltransferase [EC:2.7.7.43] [EC:2.7.7.43]	0.00	0.00	0.00	0.00	0.00	0.20
K00985	RNA-directed RNA polymerase [EC:2.7.7.48] [EC:2.7.7.48]	0.00	0.00	0.00	0.00	0.27	0.00
K01048	lysophospholipase [EC:3.1.1.5] [EC:3.1.1.5]	0.00	0.00	0.00	0.00	0.00	0.14
K01056	peptidyl-tRNA hydrolase, PTH1 family [EC:3.1.1.29] [EC:3.1.1.29]	0.00	0.00	0.00	0.00	0.64	0.12
K01058	phospholipase A1 [EC:3.1.1.32] [EC:3.1.1.32]	0.00	0.00	0.00	0.00	0.00	0.07
K01083	3-phytase [EC:3.1.3.8] [EC:3.1.3.8]	0.00	0.00	0.00	0.00	0.00	0.12
K01095	phosphatidylglycerophosphatase A [EC:3.1.3.27] [EC:3.1.3.27]	0.49	0.00	0.00	0.00	0.00	0.00
K01139	quanosine-3',5'-bis(diphosphate) 3'-pyrophosphohydrolase [EC:3.1.7.2] [EC:3	0.00	0.00	0.00	0.00	0.00	0.10
K01246	DNA-3-methyladenine glycosylase I [EC:3.2.2.20] [EC:3.2.2.20]	0.00	0.00	0.00	0.00	0.32	0.00
K01251	adenosylhomocysteinase [EC:3.3.1.1] [EC:3.3.1.1]	0.00	0.00	0.00	0.00	0.00	0.20
K01256	aminopeptidase N [EC:3.4.11.2] [EC:3.4.11.2]	0.00	0.00	0.00	0.00	0.00	0.08
K01265	methionyl aminopeptidase [EC:3.4.11.18] [EC:3.4.11.18]	0.00	0.00	0.00	0.00	0.24	0.00
K01436	aminoacylase [EC:3.5.1.14] [EC:3.5.1.14]	0.00	0.00	0.00	0.00	0.60	0.07
K01439	succinyl-diaminopimelate desuccinylase [EC:3.5.1.18] [EC:3.5.1.18]	0.00	0.00	0.00	0.00	0.17	0.06
K01443	N-acetylglucosamine-6-phosphate deacetylase [EC:3.5.1.25] [EC:3.5.1.25]	0.00	0.00	0.00	0.00	1.37	0.12
K01448	N-acetylmuramoyl-L-alanine amidase [EC:3.5.1.28] [EC:3.5.1.28]	0.00	0.00	0.00	0.00	0.29	0.00
K01484	succinylarginine dihydrolase [EC:3.5.3.23] [EC:3.5.3.23]	0.00	0.00	0.00	0.00	0.28	0.00

Data represent the number of sequence hits to each target ortholog per 10,000 reads, normalized to the gene size (in base pairs) of each specific ortholog.

Table S1. Idiomarinaceae specific KEGG orthologues in control and treatment cDNAs

K01507	inorganic pyrophosphatase [EC:3.6.1.1] [EC:3.6.1.1]	0.00	0.00	0.00	0.00	0.72	0.13
K01514	exopolyphosphatase [EC:3.6.1.11] [EC:3.6.1.11]	0.00	0.00	0.00	0.00	0.13	0.00
K01556	kynureninase [EC:3.7.1.3] [EC:3.7.1.3]	0.00	0.00	0.00	0.00	0.19	0.00
K01588	phosphoribosylaminoimidazole carboxylase catalytic subunit [EC:4.1.1.21] [EC	0.00	0.00	0.00	0.00	0.39	0.14
K01589	phosphoribosylaminoimidazole carboxylase ATPase subunit [EC:4.1.1.21] [EC:	0.00	0.00	0.00	0.00	1.05	0.00
K01598	phosphopantothenoylcysteine decarboxylase [EC:4.1.1.36]	0.00	0.00	0.00	0.00	0.00	0.17
K01610	phosphoenolpyruvate carboxykinase (ATP) [EC:4.1.1.49] [EC:4.1.1.49]	0.00	0.00	0.00	0.00	0.12	0.04
K01613	phosphatidylserine decarboxylase [EC:4.1.1.65] [EC:4.1.1.65]	0.00	0.00	0.00	0.00	0.22	0.08
K01618	glutamate decarboxylase, putative	0.00	0.00	0.00	0.00	0.34	0.19
K01620	threonine aldolase [EC:4.1.2.5] [EC:4.1.2.5]	0.00	0.00	0.00	0.00	0.00	0.07
K01658	anthranilate synthase component II [EC:4.1.3.27] [EC:4.1.3.27]	0.00	0.00	0.00	0.00	0.30	0.00
K01659	2-methylcitrate synthase [EC:2.3.3.5] [EC:2.3.3.5]	0.00	0.00	0.00	0.00	0.00	0.06
K01664	para-aminobenzoate synthetase component II [EC:2.6.1.85] [EC:2.6.1.85]	0.00	0.00	0.00	0.00	0.33	0.61
K01669	deoxyribodipyrimidine photo-lyase [EC:4.1.99.3] [EC:4.1.99.3]	0.17	0.00	0.00	0.00	0.27	0.05
K01673	carbonic anhydrase [EC:4.2.1.1] [EC:4.2.1.1]	0.00	0.00	0.00	0.00	0.30	0.00
K01679	fumarate hydratase, class II [EC:4.2.1.2] [EC:4.2.1.2]	0.00	0.00	0.00	0.00	0.00	0.10
K01682	aconitate hydratase 2 [EC:4.2.1.3] [EC:4.2.1.3]	0.00	0.00	0.00	0.00	0.37	0.16
K01714	dihydrodipicolinate synthase [EC:4.2.1.52] [EC:4.2.1.52]	0.00	0.00	0.00	0.00	0.00	0.16
K01720	2-methylcitrate dehydratase [EC:4.2.1.79] [EC:4.2.1.79]	0.16	0.00	0.00	0.00	0.00	0.00
K01745	histidine ammonia-lyase [EC:4.3.1.3] [EC:4.3.1.3]	0.00	0.00	0.00	0.00	0.37	0.00
K01749	hydroxymethylbilane synthase [EC:2.5.1.61] [EC:2.5.1.61]	0.00	0.00	0.00	0.00	0.20	0.07
K01752	L-serine dehydratase [EC:4.3.1.17] [EC:4.3.1.17]	0.00	0.00	0.00	0.00	0.14	0.00
K01759	lactoylglutathione lyase [EC:4.4.1.5] [EC:4.4.1.5]	0.00	0.00	0.00	0.00	0.39	0.28
K01760	cystathionine beta-lyase [EC:4.4.1.8] [EC:4.4.1.8]	0.00	0.00	0.00	0.00	0.39	0.00
K01763	selenocysteine lyase [EC:4.4.1.16] [EC:4.4.1.16]	0.00	0.00	0.00	0.00	0.16	0.00
K01770	2-C-methyl-D-erythritol 2,4-cyclodiphosphate synthase [EC:4.6.1.12] [EC:4.6	0.00	0.00	0.00	0.00	0.00	0.30
K01772	ferrochelatase [EC:4.99.1.1] [EC:4.99.1.1]	0.00	0.00	0.00	0.00	0.18	0.19
K01776	glutamate racemase [EC:5.1.1.3] [EC:5.1.1.3]	0.00	0.00	0.00	0.00	0.26	0.00
K01778	diaminopimelate epimerase [EC:5.1.1.7] [EC:5.1.1.7]	0.00	0.00	0.00	0.00	0.23	0.08
K01784	UDP-glucose 4-epimerase [EC:5.1.3.2] [EC:5.1.3.2]	0.00	0.00	0.00	0.00	0.00	0.07
K01800	maleylacetoacetate isomerase [EC:5.2.1.2] [EC:5.2.1.2]	0.00	0.00	0.00	0.00	0.30	0.11
K01803	triosephosphate isomerase (TIM) [EC:5.3.1.1] [EC:5.3.1.1]	0.00	0.00	0.00	0.00	0.50	0.37
K01810	glucose-6-phosphate isomerase [EC:5.3.1.9] [EC:5.3.1.9]	0.00	0.00	0.00	0.00	0.00	0.14
K01839	phosphopentomutase [EC:5.4.2.7] [EC:5.4.2.7]	0.00	0.00	0.00	0.00	0.00	0.06
K01845	glutamate-1-semialdehyde 2,1-aminomutase [EC:5.4.3.8] [EC:5.4.3.8]	0.00	0.00	0.00	0.00	0.00	0.05
K01867	tryptophanyl-tRNA synthetase [EC:6.1.1.2] [EC:6.1.1.2]	0.00	0.00	0.00	0.00	0.00	0.14
K01876	aspartyl-tRNA synthetase [EC:6.1.1.12] [EC:6.1.1.12]	0.00	0.00	0.00	0.00	0.11	0.04
K01878	glycyl-tRNA synthetase alpha chain [EC:6.1.1.14]	0.00	0.00	0.00	0.00	0.00	0.07
K01881	prolyl-tRNA synthetase [EC:6.1.1.15] [EC:6.1.1.15]	0.14	0.00	0.00	0.00	0.11	0.08
K01883	cysteinyl-tRNA synthetase [EC:6.1.1.16] [EC:6.1.1.16]	0.00	0.00	0.00	0.00	0.00	0.05
K01885	glutamyl-tRNA synthetase [EC:6.1.1.17] [EC:6.1.1.17]	0.00	0.00	0.00	0.00	0.00	0.10
K01887	arginyl-tRNA synthetase [EC:6.1.1.19] [EC:6.1.1.19]	0.00	0.00	0.00	0.00	0.11	0.12
K01890	phenylalanyl-tRNA synthetase beta chain [EC:6.1.1.20] [EC:6.1.1.20]	0.00	0.00	0.00	0.00	0.48	0.09
K01904	4-coumarate--CoA ligase [EC:6.2.1.12] [EC:6.2.1.12]	0.00	0.00	0.00	0.00	0.16	0.42
K01907	acetoacetyl-CoA synthetase [EC:6.2.1.16] [EC:6.2.1.16]	0.00	0.00	0.00	0.00	0.00	0.06
K01916	NAD+ synthase [EC:6.3.1.5] [EC:6.3.1.5]	0.00	0.00	0.00	0.00	0.27	0.00
K01918	pantoate--beta-alanine ligase [EC:6.3.2.1] [EC:6.3.2.1]	0.00	0.00	0.00	0.00	0.45	0.00
K01923	phosphoribosylaminoimidazole-succinocarboxamide synthase [EC:6.3.2.6] [EC	0.00	0.00	0.00	0.00	0.27	0.30
K01928	UDP-N-acetylmuramoylalanyl-D-glutamate--2, 6-diaminopimelate ligase [EC:6	0.00	0.00	0.00	0.00	0.13	0.00
K01929	UDP-N-acetylmuramoylalanyl-D-glutamyl-2, 6-diaminopimelate--D-alanyl-D-al	0.00	0.00	0.00	0.00	0.27	0.00
K01937	CTP synthase [EC:6.3.4.2] [EC:6.3.4.2]	0.00	0.00	0.00	0.00	0.00	0.09
K01939	adenylosuccinate synthase [EC:6.3.4.4] [EC:6.3.4.4]	0.00	0.00	0.00	0.00	0.44	0.00
K01952	phosphoribosylformylglycinamidine synthase [EC:6.3.5.3] [EC:6.3.5.3]	0.00	0.00	0.00	0.00	0.05	0.04
K01963	acetyl-CoA carboxylase carboxyl transferase subunit beta [EC:6.4.1.2] [EC:6.4	0.00	0.00	0.00	0.00	0.65	0.00
K01968	3-methylcrotonyl-CoA carboxylase alpha subunit [EC:6.4.1.4] [EC:6.4.1.4]	0.00	0.00	0.00	0.00	0.10	0.00
K01991	polysaccharide export outer membrane protein [NA]	0.00	0.00	0.00	0.00	0.44	0.20
K02010	iron(III) transport system ATP-binding protein [EC:3.6.3.30] [EC:3.6.3.30]	0.00	0.00	0.00	0.00	0.18	0.00
K02011	iron(III) transport system permease protein [NA]	0.00	0.00	0.00	0.00	0.59	0.00
K02012	iron(III) transport system substrate-binding protein [NA]	0.00	0.00	0.00	0.00	0.19	0.00
K02034	peptide/nickel transport system permease protein [NA]	0.00	0.00	0.00	0.00	0.00	0.09
K02035	peptide/nickel transport system substrate-binding protein [NA]	0.00	0.00	0.00	0.00	0.12	0.09
K02037	phosphate transport system permease protein [NA]	0.00	0.00	0.00	0.00	0.34	0.34
K02045	sulfate transport system ATP-binding protein [EC:3.6.3.25] [EC:3.6.3.25]	0.00	0.00	0.00	0.00	0.00	0.07
K02066	putative ABC transport system permease protein [NA]	0.00	0.00	0.00	0.00	0.49	0.90
K02108	F-type H+-transporting ATPase subunit a [EC:3.6.3.14] [EC:3.6.3.14]	0.00	0.00	0.21	0.00	0.72	0.62
K02109	F-type H+-transporting ATPase subunit b [EC:3.6.3.14] [EC:3.6.3.14]	0.00	0.00	0.00	0.00	1.22	0.00
K02111	F-type H+-transporting ATPase subunit alpha [EC:3.6.3.14] [EC:3.6.3.14]	0.00	0.00	0.00	0.00	0.25	0.09
K02113	F-type H+-transporting ATPase subunit delta [EC:3.6.3.14] [EC:3.6.3.14]	0.00	0.00	0.00	0.00	1.08	0.66
K02114	F-type H+-transporting ATPase subunit epsilon [EC:3.6.3.14] [EC:3.6.3.14]	0.00	0.00	0.00	0.00	0.00	0.17
K02115	F-type H+-transporting ATPase subunit gamma [EC:3.6.3.14] [EC:3.6.3.14]	0.00	0.00	0.00	0.00	0.00	0.16
K02116	ATP synthase protein I [NA]	0.00	0.00	0.00	0.00	0.50	0.00
K02195	heme exporter membrane protein CcmC [NA]	0.00	0.00	0.00	0.00	0.26	0.38
K02196	cytochrome c-type biogenesis protein CcmD [NA]	0.00	0.00	0.00	0.00	0.00	0.30
K02314	replicative DNA helicase [EC:3.6.1.-] [EC:3.6.1.-]	0.00	0.00	0.00	0.00	0.14	0.10
K02339	DNA polymerase III subunit chi [EC:2.7.7.7] [EC:2.7.7.7]	0.00	0.00	0.00	0.00	0.00	0.15
K02340	DNA polymerase III subunit delta [EC:2.7.7.7] [EC:2.7.7.7]	0.00	0.00	0.00	0.00	0.18	0.00
K02413	flagellar FliJ protein [NA]	0.00	0.00	0.00	0.00	0.00	0.16
K02417	flagellar motor switch protein FliN/FliY [NA]	0.00	0.00	0.00	0.00	0.42	0.00
K02419	flagellar biosynthetic protein FliP [NA]	0.00	0.00	0.00	0.00	0.26	0.00
K02420	flagellar biosynthetic protein FliQ [NA]	0.00	0.00	0.00	0.00	0.72	0.00
K02421	flagellar biosynthetic protein FliR [NA]	0.00	0.00	0.00	0.00	0.00	0.27
K02427	cell division protein methyltransferase FtsJ [EC:2.1.1.-] [EC:2.1.1.-]	0.00	0.00	0.00	0.00	0.31	0.11
K02483	two-component system, OmpR family, response regulator [NA]	0.00	0.00	0.00	0.00	0.00	0.10
K02484	two-component system, OmpR family, sensorinase [EC:2.7.13.3] [EC:2.7.13.3	0.00	0.00	0.00	0.00	0.00	0.10

Data represent the number of sequence hits to each target ortholog per 10,000 reads, normalized to the gene size (in base pairs) of each specific ortholog.

Table S1. Idiomarinaceae specific KEGG orthologues in control and treatment cDNAs

K02495	oxygen-independent coproporphyrinogen III oxidase [EC:1.3.99.22] [EC:1.3.9	0.19	0.00	0.00	0.00	0.00	0.00
K02505	protein transport protein HofC [NA]	0.00	0.00	0.00	0.00	0.00	0.06
K02527	3-deoxy-D-manno-octulosonic-acid transferase [EC:2.-.-.-] [EC:2.-.-.-]	0.00	0.00	0.00	0.00	0.10	0.03
K02535	UDP-3-O-[3-hydroxymyristoyl] N-acetylglucosamine deacetylase [EC:3.5.1.-]	0.00	0.00	0.00	0.00	1.45	1.00
K02563	UDP-N-acetylglucosamine--N-acetylmuramyl-(pentapeptide) pyrophosphoryl-u	0.00	0.00	0.00	0.00	0.18	0.00
K02584	Nif-specific regulatory protein [NA]	0.00	0.00	0.00	0.00	0.35	0.00
K02601	transcriptional antiterminator NusG [NA]	0.00	0.00	0.00	0.00	1.42	0.65
K02656	type IV pilus assembly protein PilF [NA]	0.00	0.00	0.00	0.00	0.00	0.08
K02667	two-component system, NtrC family, response regulator PilR [NA]	0.00	0.00	0.00	0.00	0.15	0.00
K02687	ribosomal protein L11 methyltransferase [EC:2.1.1.-] [EC:2.1.1.-]	0.00	0.00	0.00	0.00	0.22	0.00
K02806	PTS system, nitrogen regulatory IIA component [EC:2.7.1.69] [EC:2.7.1.69]	0.00	0.00	0.00	0.00	0.41	0.00
K02834	ribosome-binding factor A [NA]	0.00	0.00	0.00	0.00	0.00	0.18
K02860	16S rRNA processing protein RimM [NA]	0.00	0.00	0.00	0.00	1.10	0.67
K02863	large subunit ribosomal protein L1 [NA]	0.34	0.00	0.00	0.00	0.00	0.20
K02871	large subunit ribosomal protein L13 [NA]	0.00	0.00	0.00	0.00	0.00	0.16
K02876	large subunit ribosomal protein L15 [NA]	0.00	0.00	0.00	0.00	1.76	1.30
K02881	large subunit ribosomal protein L18 [NA]	0.00	0.00	0.00	0.00	1.09	0.20
K02886	large subunit ribosomal protein L2 [NA]	0.00	0.00	0.20	0.00	0.46	0.17
K02887	large subunit ribosomal protein L20 [NA]	0.00	0.00	0.00	0.00	0.55	0.00
K02888	large subunit ribosomal protein L21 [NA]	0.00	0.00	0.00	0.00	0.00	0.46
K02890	large subunit ribosomal protein L22 [NA]	0.00	0.00	0.00	0.00	1.15	0.00
K02892	large subunit ribosomal protein L23 [NA]	0.00	0.00	0.00	0.00	0.64	0.23
K02897	large subunit ribosomal protein L25 [NA]	0.00	0.00	0.25	0.00	1.44	1.70
K02906	large subunit ribosomal protein L3 [NA]	0.00	0.00	0.26	0.00	1.20	0.77
K02911	large subunit ribosomal protein L32 [NA]	1.41	0.00	0.00	0.00	2.27	2.52
K02931	large subunit ribosomal protein L5 [NA]	0.00	0.00	0.00	0.00	0.71	0.39
K02933	large subunit ribosomal protein L6 [NA]	0.89	0.00	0.00	0.00	0.36	1.59
K02935	large subunit ribosomal protein L7/L12 [NA]	0.64	0.00	0.00	0.00	0.52	0.94
K02939	large subunit ribosomal protein L9 [NA]	0.00	0.00	0.00	0.00	0.43	0.15
K02945	small subunit ribosomal protein S1 [NA]	0.00	0.00	0.00	0.00	0.11	0.13
K02952	small subunit ribosomal protein S13 [NA]	0.00	0.00	0.47	0.00	2.69	1.19
K02954	small subunit ribosomal protein S14 [NA]	0.00	0.00	0.00	0.00	0.00	1.16
K02956	small subunit ribosomal protein S15 [NA]	0.00	0.00	0.00	0.00	0.72	0.00
K02959	small subunit ribosomal protein S16 [NA]	0.00	0.00	0.00	0.00	0.78	0.00
K02961	small subunit ribosomal protein S17 [NA]	0.00	0.00	0.00	0.00	0.74	0.26
K02967	small subunit ribosomal protein S2 [NA]	0.00	0.00	0.00	0.00	0.26	0.68
K02968	small subunit ribosomal protein S20 [NA]	0.00	0.00	0.00	0.00	0.00	0.27
K02986	small subunit ribosomal protein S4 [NA]	0.00	0.00	0.00	0.00	0.31	0.00
K02988	small subunit ribosomal protein S5 [NA]	0.00	0.00	0.00	0.00	2.30	0.57
K02990	small subunit ribosomal protein S6 [NA]	0.00	0.00	0.00	0.00	0.00	1.01
K02994	small subunit ribosomal protein S8 [NA]	0.00	0.00	0.00	0.00	0.98	0.36
K02996	small subunit ribosomal protein S9 [NA]	0.00	0.00	0.00	0.00	0.98	0.00
K03040	DNA-directed RNA polymerase subunit alpha [EC:2.7.7.6] [EC:2.7.7.6]	0.00	0.00	0.00	0.00	0.19	0.14
K03043	DNA-directed RNA polymerase subunit beta [EC:2.7.7.6] [EC:2.7.7.6]	0.00	0.00	0.00	0.00	0.47	0.09
K03060	DNA-directed RNA polymerase subunit omega [EC:2.7.7.6] [EC:2.7.7.6]	0.00	0.00	0.00	0.00	1.43	0.00
K03070	preprotein translocase SecA subunit [NA]	0.00	0.00	0.00	0.00	0.35	0.00
K03073	preprotein translocase SecE subunit [NA]	0.00	0.00	0.00	0.00	0.51	0.75
K03074	preprotein translocase SecF subunit [NA]	0.00	0.00	0.00	0.00	0.00	0.15
K03075	preprotein translocase SecG subunit [NA]	0.00	0.00	0.00	0.00	1.69	0.62
K03076	preprotein translocase SecY subunit [NA]	0.00	0.00	0.00	0.00	0.00	0.21
K03088	RNA polymerase sigma-70 factor, ECF subfamily [NA]	0.40	0.00	0.00	0.00	0.64	1.53
K03106	signal recognition particle, subunit SRP54 [NA]	0.00	0.00	0.00	0.00	0.28	0.00
K03118	sec-independent protein translocase protein TatC [NA]	0.00	0.00	0.00	0.00	0.00	0.09
K03149	thiamine biosynthesis ThiG [NA]	0.00	0.00	0.00	0.00	0.25	0.00
K03150	thiamine biosynthesis ThiH [NA]	0.00	0.00	0.00	0.00	0.17	0.00
K03151	thiamine biosynthesis protein ThiI [NA]	0.00	0.00	0.00	0.00	0.00	0.10
K03177	tRNA pseudouridine synthase B [EC:5.4.99.12] [EC:5.4.99.12]	0.00	0.00	0.00	0.00	0.00	0.15
K03179	4-hydroxybenzoate octaprenyltransferase [EC:2.5.1.-] [EC:2.5.1.-]	0.00	0.00	0.00	0.00	0.00	0.25
K03182	3-octaprenyl-4-hydroxybenzoate carboxy-lyase UbiD [EC:4.1.1.-] [EC:4.1.1.-]	0.00	0.00	0.00	0.00	0.26	0.05
K03183	ubiquinone/menaquinone biosynthesis methyltransferase [EC:2.1.1.-] [EC:2.1	0.00	0.00	0.00	0.00	0.51	0.09
K03184	2-octaprenyl-3-methyl-6-methoxy-1,4-benzoquinol hydroxylase [EC:1.14.13.-	0.00	0.00	0.00	0.00	0.00	0.06
K03186	3-octaprenyl-4-hydroxybenzoate carboxy-lyase UbiX [EC:4.1.1.-] [EC:4.1.1.-]	0.00	0.00	0.00	0.00	0.31	0.00
K03210	preprotein translocase YajC subunit [NA]	0.00	0.00	0.00	0.00	0.58	0.42
K03215	RNA methyltransferase, TrmA family [EC:2.1.1.-] [EC:2.1.1.-]	0.00	0.00	0.00	0.00	0.42	0.16
K03216	RNA methyltransferase, TrmH family, group 2 [EC:2.1.1.-] [EC:2.1.1.-]	0.00	0.00	0.00	0.00	0.00	0.31
K03217	preprotein translocase YidC subunit [NA]	0.00	0.00	0.00	0.00	0.24	0.09
K03218	RNA methyltransferase, TrmH family [EC:2.1.1.-] [EC:2.1.1.-]	0.00	0.00	0.00	0.00	1.02	0.00
K03269	UDP-2,3-diacylglucosamine hydrolase [EC:3.6.1.-] [EC:3.6.1.-]	0.00	0.00	0.00	0.00	0.27	0.10
K03284	metal ion transporter, MIT family [NA]	0.00	0.00	0.00	0.00	0.20	0.00
K03295	cation efflux system protein, CDF family [NA]	0.00	0.00	0.00	0.00	0.43	0.00
K03305	proton-dependent oligopeptide transporter, POT family [NA]	0.00	0.00	0.00	0.00	0.37	0.04
K03308	neurotransmitter:Na+ symporter, NSS family [NA]	0.00	0.00	0.12	0.00	0.55	0.10
K03315	Na+:H+ antiporter, NhaC family [NA]	0.00	0.00	0.00	0.00	0.39	0.10
K03322	metal ion transporter, Nramp family [NA]	0.00	0.00	0.00	0.00	0.16	0.12
K03325	arsenite transporter, ACR3 family [NA]	0.00	0.00	0.00	0.00	0.18	0.00
K03327	multidrug resistance protein, MATE family [NA]	0.00	0.00	0.00	0.00	0.14	0.31
K03415	two-component system, chemotaxis family, response regulator CheV [NA]	0.00	0.00	0.00	0.00	0.00	0.23
K03424	Mg-dependent DNase [EC:3.1.21.-] [EC:3.1.21.-]	0.00	0.00	0.00	0.00	0.48	0.00
K03426	NAD+ diphosphatase [EC:3.6.1.22] [EC:3.6.1.22]	0.00	0.00	0.00	0.00	0.48	0.09
K03431	phosphoglucosamine mutase [EC:5.4.2.10] [EC:5.4.2.10]	0.00	0.00	0.00	0.00	0.28	0.00
K03451	betaine/carnitine transporter, BCCT family [NA]	0.00	0.00	0.00	0.00	0.11	0.04
K03469	ribonuclease HI [EC:3.1.26.4] [EC:3.1.26.4]	0.00	0.00	0.00	0.00	0.81	0.00
K03495	glucose inhibited division protein A [NA]	0.00	0.00	0.00	0.00	0.20	0.07
K03517	quinolinate synthase [NA]	0.00	0.00	0.00	0.00	0.00	0.14

Data represent the number of sequence hits to each target ortholog per 10,000 reads, normalized to the gene size (in base pairs) of each specific ortholog.

Table S1. Idiomarinaceae specific KEGG orthologues in control and treatment cDNAs

K03525	type III pantothenateinase [EC:2.7.1.33] [EC:2.7.1.33]	0.00	0.00	0.00	0.00	0.00	0.09
K03536	ribonuclease P protein component [EC:3.1.26.5] [EC:3.1.26.5]	0.00	0.00	0.00	0.00	0.00	0.18
K03543	multidrug resistance protein A [NA]	0.00	0.00	0.00	0.00	0.51	0.00
K03545	trigger factor [NA]	0.00	0.00	0.00	0.00	0.15	0.00
K03551	holliday junction DNA helicase RuvB [NA]	0.00	0.00	0.00	0.00	0.19	0.00
K03553	recombination protein RecA [NA]	0.00	0.00	0.00	0.00	0.00	0.06
K03555	DNA mismatch repair protein MutS [NA]	0.00	0.00	0.00	0.00	0.08	0.03
K03563	carbon storage regulator [NA]	0.00	0.00	0.00	0.00	0.00	1.89
K03572	DNA mismatch repair protein MutL [NA]	0.00	0.00	0.00	0.00	0.00	0.04
K03584	DNA repair protein RecO (recombination protein O) [NA]	0.00	0.00	0.00	0.00	0.27	0.00
K03585	membrane fusion protein [NA]	0.00	0.00	0.00	0.00	0.00	0.12
K03591	cell division protein FtsN [NA]	0.00	0.00	0.00	0.00	0.00	0.06
K03599	stringent starvation protein A [NA]	0.00	0.00	0.00	0.00	0.61	0.23
K03611	disulfide bond formation protein DsbB [NA]	0.00	0.00	0.00	0.00	0.00	0.39
K03625	N utilization substance protein B [NA]	0.57	1.12	0.00	0.00	0.00	0.00
K03631	DNA repair protein RecN (Recombination protein N) [NA]	0.14	0.00	0.00	0.00	0.00	0.00
K03646	colicin import membrane protein [NA]	0.00	0.00	0.00	0.00	0.24	0.00
K03648	uracil-DNA glycosylase [EC:3.2.2.-] [EC:3.2.2.-]	0.00	0.00	0.00	0.00	0.00	0.65
K03657	DNA helicase II / ATP-dependent DNA helicase PcrA [EC:3.6.1.-] [EC:3.6.1.-]	0.00	0.00	0.00	0.00	0.18	0.00
K03664	SsrA-binding protein [NA]	0.00	0.00	0.00	0.00	0.00	0.14
K03668	heat shock protein HslJ [NA]	0.18	0.00	0.00	0.00	0.00	0.00
K03683	ribonuclease T [EC:3.1.13.-] [EC:3.1.13.-]	0.00	0.00	0.00	0.00	0.00	0.11
K03685	ribonuclease III [EC:3.1.26.3] [EC:3.1.26.3]	0.00	0.00	0.00	0.00	0.00	0.41
K03688	ubiquinone biosynthesis protein [NA]	0.00	0.00	0.00	0.00	0.34	0.00
K03701	excinuclease ABC subunit A [NA]	0.00	0.00	0.00	0.00	0.00	0.02
K03722	ATP-dependent DNA helicase DinG [EC:3.6.1.-] [EC:3.6.1.-]	0.00	0.00	0.00	0.00	0.48	0.04
K03723	transcription-repair coupling factor (superfamily II helicase) [EC:3.6.1.-] [EC:3	0.00	0.00	0.00	0.00	0.06	0.04
K03733	integrase/recombinase XerC [NA]	0.00	0.00	0.00	0.00	0.00	0.16
K03770	peptidyl-prolyl cis-trans isomerase D [EC:5.2.1.8] [EC:5.2.1.8]	0.00	0.00	0.00	0.00	0.20	0.08
K03775	FKBP-type peptidyl-prolyl cis-trans isomerase SlyD [EC:5.2.1.8] [EC:5.2.1.8]	0.00	0.00	0.00	0.00	0.00	0.14
K03789	ribosomal-protein-alanine N-acetyltransferase [EC:2.3.1.128] [EC:2.3.1.128]	0.00	0.00	0.00	0.00	0.88	0.32
K03801	lipoyl(octanoyl) transferase [EC:2.3.1.181] [EC:2.3.1.181]	0.00	0.00	0.00	0.00	0.00	0.10
K03811	nicotinamide mononucleotide transporter [NA]	0.00	0.00	0.00	0.00	0.31	0.11
K04042	bifunctional protein GlmU [EC:2.3.1.157 2.7.7.23] [EC:2.3.1.157 2.7.7.23]	0.00	0.00	0.00	0.00	0.52	0.29
K04043	molecular chaperone DnaK [NA]	0.00	0.00	0.00	0.00	0.00	0.04
K04075	cell cycle protein MesJ [EC:6.3.4.-] [EC:6.3.4.-]	0.00	0.00	0.00	0.00	0.00	0.05
K04567	lysyl-tRNA synthetase, class II [EC:6.1.1.6] [EC:6.1.1.6]	0.00	0.00	0.00	0.00	0.13	0.00
K04568	lysyl-tRNA synthetase, class II [EC:6.1.1.6] [EC:6.1.1.6]	0.00	0.00	0.00	0.00	0.00	0.14
K05350	beta-glucosidase [EC:3.2.1.21] [EC:3.2.1.21]	0.24	0.00	0.00	0.00	0.00	0.00
K05526	succinylglutamate desuccinylase [EC:3.5.1.96] [EC:3.5.1.96]	0.00	0.00	0.00	0.00	0.19	0.00
K05540	tRNA-dihydrouridine synthase B [EC:1.-.-.-] [EC:1.-.-.-]	0.00	0.00	0.00	0.00	0.20	0.44
K05577	NADH dehydrogenase I subunit 5 [EC:1.6.5.3] [EC:1.6.5.3]	0.00	0.00	0.00	0.00	0.15	0.05
K05590	ATP-dependent RNA helicase SrmB [EC:2.7.7.-] [EC:2.7.7.-]	0.19	0.00	0.00	0.00	0.16	0.11
K05592	ATP-dependent RNA helicase DeaD [NA]	0.27	0.00	0.00	0.00	0.43	0.04
K05779	putative thiamine transport system ATP-binding protein [NA]	0.00	0.00	0.00	0.00	0.00	0.10
K05786	chloramphenicol-sensitive protein RarD [NA]	0.00	0.00	0.00	0.00	0.21	0.00
K05802	potassium efflux system proteinefA [NA]	0.00	0.00	0.00	0.00	0.00	0.42
K05896	segregation and condensation protein A [NA]	0.00	0.26	0.00	0.00	0.21	0.00
K06024	segregation and condensation protein B [NA]	0.00	0.00	0.00	0.00	0.00	0.12
K06158	ATP-binding cassette, sub-family F, member 3 [NA]	0.00	0.24	0.00	0.00	0.00	0.00
K06169	tRNA-(ms[2]io[6]A)-hydroxylase [EC:1.-.-.-] [2]io[6]A)-hydroxylase [EC:1.-.-	0.00	0.00	0.00	0.00	0.00	0.09
K06175	tRNA pseudouridine synthase C [EC:5.4.99.12] [EC:5.4.99.12]	0.00	0.00	0.00	0.00	0.00	0.10
K06181	ribosomal large subunit pseudouridine synthase E [EC:5.4.99.12] [EC:5.4.99.1	0.00	0.00	0.00	0.00	0.00	0.37
K06187	recombination protein RecR [NA]	0.00	0.00	0.00	0.00	0.32	0.23
K06190	intracellular septation protein [NA]	0.00	0.00	0.00	0.00	0.00	0.13
K06213	magnesium transporter [NA]	0.00	0.00	0.00	0.00	0.00	0.05
K06350	antagonist ofipI [NA]	0.00	0.00	0.00	0.00	0.22	0.08
K07320	putative adenine-specific DNA-methyltransferase [EC:2.1.1.72] [EC:2.1.1.72]	0.00	0.00	0.00	0.00	0.00	0.07
K07397	putative redox protein [NA]	0.00	0.00	0.00	0.00	0.00	0.16
K07478	putative ATPase [NA]	0.00	0.00	0.00	0.00	0.37	0.00
K07640	two-component system, OmpR family, sensor histidineinase CpxA [EC:2.7.13.:	0.00	0.00	0.00	0.00	0.15	0.05
K07657	two-component system, OmpR family, phosphate regulon response regulator F	0.00	0.00	0.00	0.00	1.11	1.13
K07666	two-component system, OmpR family, response regulator QseB [NA]	0.00	0.00	0.00	0.00	0.27	0.00
K07673	two-component system, NarL family, nitrate/nitrite sensor histidineinase NarX	0.00	0.00	0.00	0.00	0.11	0.04
K07678	two-component system, NarL family, sensor histidineinase BarA [EC:2.7.13.3]	0.00	0.00	0.00	0.00	0.11	0.00
K07679	two-component system, NarL family, sensor histidineinase EvgS [EC:2.7.13.3]	0.00	0.00	0.00	0.00	0.11	0.00
K07712	two-component system, NtrC family, nitrogen regulation response regulator GI	0.00	0.21	0.00	0.00	0.00	0.00
K07799	putative multidrug efflux transporter MdtA [NA]	0.00	0.00	0.00	0.00	0.00	0.09
K08485	phosphocarrier protein NPr [NA]	0.00	0.00	0.00	0.00	0.00	0.77
K09458	3-oxoacyl-[acyl-carrier-protein] synthase II [EC:2.3.1.179] [acyl-carrier-prote	0.00	0.00	0.00	0.00	0.00	0.06
K09687	antibiotic transport system ATP-binding protein [NA]	0.00	0.00	0.00	0.00	0.00	0.17
K09696	sodium transport system permease protein [NA]	0.00	0.00	0.00	0.00	0.16	0.00
K09808	lipoprotein-releasing system permease protein [NA]	0.00	0.00	0.00	0.00	0.47	0.06
K09810	lipoprotein-releasing system ATP-binding protein [EC:3.6.3.-] [EC:3.6.3.-]	0.00	0.00	0.00	0.00	0.00	0.10
K09811	cell division transport system permease protein [NA]	0.00	0.00	0.00	0.00	0.19	0.00
K10563	formamidopyrimidine-DNA glycosylase [EC:3.2.2.23 4.2.99.18] [EC:3.2.2.23 4	0.00	0.00	0.00	0.00	0.00	0.09
K10804	acyl-CoA thioesterase I [EC:3.1.2.-] [EC:3.1.2.-]	0.00	0.00	0.00	0.00	0.00	0.10

Data represent the number of sequence hits to each target ortholog per 10,000 reads, normalized to the gene size (in base pairs) of each specific ortholog.

Table S2. Alteromonadaceae specific KEGG orthologues in control and treatment cDNAs

Definition	con_2hrs	con_12 hrs	con_27 hrs	DOM_2 hrs	DOM_12 hrs	DOM_27 hrs
large subunit ribosomal protein L18 [NA]	0.65	0.64	1.37	0.56	12.60	2.13
16S rRNA processing protein RimM [NA]	0.91	1.79	2.56	0.39	9.18	4.47
tRNA (guanine-N1-)-methyltransferase [EC:2.1.1.31] [EC:2.1.1.31]	0.31	1.22	1.75	1.07	7.03	8.43
large subunit ribosomal protein L4 [NA]	0.39	0.38	2.48	1.69	6.95	2.33
NAD(P) transhydrogenase subunit beta [EC:1.6.1.2] [EC:1.6.1.2]	0.00	0.00	0.00	0.45	6.79	5.22
large subunit ribosomal protein L15 [NA]	0.00	1.61	1.92	0.94	6.62	2.93
NAD(P) transhydrogenase subunit alpha [EC:1.6.1.2]	0.00	0.32	0.00	0.85	6.38	6.86
small subunit ribosomal protein S13 [NA]	0.67	0.00	1.87	1.14	5.92	3.77
large subunit ribosomal protein L30 [NA]	1.34	0.00	0.00	1.15	5.39	0.39
large subunit ribosomal protein L29 [NA]	0.00	0.00	0.87	0.00	5.05	1.11
large subunit ribosomal protein L6 [NA]	0.00	0.44	0.31	0.00	5.02	0.53
small subunit ribosomal protein S8 [NA]	1.22	0.00	1.28	0.52	4.88	1.44
large subunit ribosomal protein L23 [NA]	1.00	0.00	0.70	0.00	4.82	0.59
isocitrate lyase [EC:4.1.3.1] [EC:4.1.3.1]	0.45	0.29	0.00	1.27	4.78	5.73
iron complex outermembrane recepter protein [NA]	0.22	0.21	0.30	1.21	4.55	2.48
methyl-accepting chemotaxis protein [NA]	1.03	0.00	0.72	0.88	4.54	4.10
large subunit ribosomal protein L3 [NA]	0.00	0.00	0.78	0.64	4.50	1.22
malate dehydrogenase [EC:1.1.1.37] [EC:1.1.1.37]	0.00	0.25	0.00	0.22	4.28	5.26
F-type H+-transporting ATPase subunit delta [EC:3.6.3.14] [EC:3.6.3.14]	0.00	0.00	0.00	2.14	4.03	0.00
3-hydroxyacyl-CoA dehydrogenase [EC:1.1.1.35]	0.00	0.14	0.20	0.84	3.94	2.70
glutamate decarboxylase [EC:4.1.1.15] [EC:4.1.1.15]	0.00	0.00	0.00	0.00	3.74	1.38
large subunit ribosomal protein L7/L12 [NA]	0.00	0.62	2.22	0.00	3.56	3.00
small subunit ribosomal protein S3 [NA]	0.00	0.00	0.00	0.48	3.55	0.99
F-type H+-transporting ATPase subunit b [EC:3.6.3.14] [EC:3.6.3.14]	0.00	0.61	1.74	1.06	3.50	1.47
large subunit ribosomal protein L32 [NA]	0.00	0.00	0.98	0.00	3.41	5.02
large subunit ribosomal protein L25 [NA]	0.00	0.00	0.79	1.61	3.33	3.01
aromatic-amino-acid transaminase [EC:2.6.1.57] [EC:2.6.1.57]	0.00	0.00	0.00	0.00	3.29	0.81
acyl-CoA dehydrogenase [EC:1.3.99.-] [EC:1.3.99.-]	0.00	0.09	0.20	0.50	3.25	4.11
small subunit ribosomal protein S9 [NA]	1.20	1.17	0.00	0.00	2.89	1.42
two-component system, chemotaxis family, response regulator CheY [NA]	0.00	0.00	0.00	0.48	2.70	1.33
acyl-CoA dehydrogenase [EC:1.3.99.3] [EC:1.3.99.3]	0.00	0.00	0.18	0.22	2.50	2.84
acyl-CoA dehydrogenase-like protein	0.41	0.00	0.00	0.70	2.45	0.72
large subunit ribosomal protein L24 [NA]	0.00	0.74	0.00	1.30	2.44	0.67
cb-type cytochrome c oxidase subunit III [EC:1.9.3.1] [EC:1.9.3.1]	0.59	0.00	0.41	0.50	2.36	0.87
pyruvate,water dikinase [EC:2.7.9.2] [EC:2.7.9.2]	0.00	0.00	0.00	0.17	2.33	1.45
small subunit ribosomal protein S16 [NA]	0.00	0.00	0.00	0.00	2.30	0.57
small subunit ribosomal protein S5 [NA]	0.00	1.40	0.33	0.41	2.30	0.57
preprotein translocase SecE subunit [NA]	0.00	0.00	0.49	0.00	2.27	0.21
small subunit ribosomal protein S2 [NA]	0.00	0.35	0.00	0.62	2.05	0.97
phosphoadenosine phosphosulfate reductase [EC:1.8.4.8] [EC:1.8.4.8]	0.00	0.31	0.00	0.00	2.05	0.75
two-component system, NtrC family, response regulator YfhA [NA]	0.00	0.52	0.25	0.30	1.99	2.21
acetolactate synthase I/II/III large subunit [EC:2.2.1.6] [EC:2.2.1.6]	0.00	0.00	0.00	0.36	1.93	1.72
acetolactate synthase I/III small subunit [EC:2.2.1.6] [EC:2.2.1.6]	0.00	0.00	0.00	0.41	1.93	1.13
small subunit ribosomal protein S6 [NA]	0.00	0.00	0.00	1.53	1.91	0.71
F-type H+-transporting ATPase subunit alpha [EC:3.6.3.14] [EC:3.6.3.14]	0.00	0.00	0.11	0.26	1.86	0.55
tRNA-dihydrouridine synthase B [EC:1.-.-.-] [EC:1.-.-.-]	0.00	0.22	0.32	0.20	1.82	2.22
large subunit ribosomal protein L13 [NA]	0.00	0.00	0.00	0.00	1.79	1.32
nicotinamide mononucleotide transporter [NA]	0.00	0.00	0.00	0.00	1.78	0.00
ribonuclease III [EC:3.1.26.3] [EC:3.1.26.3]	0.00	0.00	0.00	0.00	1.67	0.62
malate synthase [EC:2.3.3.9] [EC:2.3.3.9]	0.11	0.00	0.15	0.18	1.65	3.14
RNA-directed DNA polymerase [EC:2.7.7.49] [EC:2.7.7.49]	0.00	0.33	0.00	0.00	1.63	1.10
glutamine synthetase [EC:6.3.1.2] [EC:6.3.1.2]	0.00	0.00	0.00	0.14	1.62	1.45
cb-type cytochrome c oxidase subunit II [EC:1.9.3.1] [EC:1.9.3.1]	0.00	0.00	0.00	0.00	1.58	0.11
F-type H+-transporting ATPase subunit beta [EC:3.6.3.14] [EC:3.6.3.14]	0.00	0.00	0.12	0.15	1.52	0.61
small subunit ribosomal protein S17 [NA]	0.00	0.00	0.00	0.80	1.49	0.27
glutamate synthase (NADPH/NADH) small chain [EC:1.4.1.13 1.4.1.14] [EC:1.4.1.1	0.00	0.16	0.12	0.00	1.48	1.94
sulfite reductase (NADPH) hemoprotein beta-component [EC:1.8.1.2] [EC:1.8.1.2]	0.00	0.40	0.29	0.00	1.44	0.78
F-type H+-transporting ATPase subunit epsilon [EC:3.6.3.14] [EC:3.6.3.14]	0.00	0.56	0.00	0.50	1.39	2.91
two-component system, OmpR family, phosphate regulon response regulator PhoB [	0.00	0.00	0.00	0.00	1.39	1.94
sulfite reductase (NADPH) flavoprotein alpha-component [EC:1.8.1.2] [NADPH] flav	0.00	0.00	0.00	0.00	1.36	0.77
glucan endo-1,3-beta-D-glucosidase [EC:3.2.1.39] [EC:3.2.1.39]	0.00	0.00	0.00	0.00	1.31	0.00
sec-independent protein translocase protein TatB [NA]	0.00	0.00	0.00	0.00	1.28	0.00
glyceraldehyde 3-phosphate dehydrogenase [EC:1.2.1.12] [EC:1.2.1.12]	0.00	0.00	0.00	0.00	1.25	1.38
6-phosphogluconate dehydrogenase [EC:1.1.1.44] [EC:1.1.1.44]	0.00	0.30	0.00	0.00	1.23	0.36
phosphoenolpyruvate carboxykinase (ATP) [EC:4.1.1.49] [EC:4.1.1.49]	0.00	0.00	0.00	0.26	1.22	1.62
UDP-N-acetylglucosamine 1-carboxyvinyltransferase [EC:2.5.1.7] [EC:2.5.1.7]	0.00	0.00	0.00	0.00	1.21	0.22
ubiquinol-cytochrome c reductase iron-sulfur subunit [EC:1.10.2.2] [EC:1.10.2.2]	0.00	0.00	0.26	0.32	1.20	0.88
thiamine-monophosphateinase [EC:2.7.4.16] [EC:2.7.4.16]	0.00	0.00	0.00	0.00	1.19	0.07
ubiquinol-cytochrome c reductase cytochrome c1 subunit [EC:1.10.2.2] [EC:1.10.2.	0.00	0.00	0.00	0.00	1.19	0.21
large subunit ribosomal protein L2 [NA]	0.00	0.57	0.61	0.00	1.16	0.43
F-type H+-transporting ATPase subunit a [EC:3.6.3.14] [EC:3.6.3.14]	0.00	0.00	0.00	0.49	1.15	1.18
myo-inositol-1(or 4)-monophosphatase [EC:3.1.3.25] [EC:3.1.3.25]	0.00	0.34	0.24	0.30	1.11	0.72
homoserine dehydrogenase [EC:1.1.1.3]	0.00	0.00	0.07	0.17	1.11	0.90
succinate dehydrogenase hydrophobic membrane anchor protein [EC:1.3.99.1] [EC:	0.00	0.00	0.48	0.00	1.10	1.83
phosphoserine aminotransferase [EC:2.6.1.52] [EC:2.6.1.52]	0.00	0.00	0.00	0.00	1.10	0.00
putative thioredoxin [NA]	0.00	0.00	0.00	0.00	1.10	0.00
triosephosphate isomerase (TIM) [EC:5.3.1.1] [EC:5.3.1.1]	0.00	0.00	0.00	0.29	1.09	1.11
3-isopropylmalate/(R)-2-methylmalate dehydratase large subunit [EC:4.2.1.33] [EC	0.00	0.00	0.24	0.29	1.09	0.91
aspartate-semialdehyde dehydrogenase [EC:1.2.1.11] [EC:1.2.1.11]	0.22	0.00	0.00	0.00	1.07	0.66
glutamate synthase (NADPH/NADH) large chain [EC:1.4.1.13 1.4.1.14] [EC:1.4.1.1	0.00	0.16	0.08	0.23	1.04	0.69

Data represent the number of sequence hits to each target ortholog per 10,000 reads, normalized to the gene size (in base pairs) of each specific ortholog.

Table S2. Alteromonadaceae specific KEGG orthologues in control and treatment cDNAs

large subunit ribosomal protein L14 [NA]	0.00	0.00	0.00	0.56	1.04	1.53
ketol-acid reductoisomerase [EC:1.1.1.86] [EC:1.1.1.86]	0.00	0.16	0.11	0.00	1.03	2.38
phosphoglucomutase [EC:5.4.2.2] [EC:5.4.2.2]	0.00	0.00	0.00	0.00	1.02	0.94
homoserineinase [EC:2.7.1.39] [EC:2.7.1.39]	0.25	0.24	0.17	0.00	1.01	1.04
phosphate transport system substrate-binding protein [NA]	0.00	0.00	0.00	0.00	1.01	0.74
flagellum-specific ATP synthase [EC:3.6.3.14] [EC:3.6.3.14]	0.00	0.17	0.00	0.00	1.00	0.37
large subunit ribosomal protein L35 [NA]	0.00	0.00	0.00	0.00	0.98	1.08
cell division protein methyltransferase FtsJ [EC:2.1.1.-] [EC:2.1.1.-]	0.00	0.00	0.00	0.00	0.98	0.00
aconitate hydratase 2 [EC:4.2.1.3] [EC:4.2.1.3]	0.09	0.08	0.06	0.30	0.97	0.56
dihydrodipicolinate reductase [EC:1.3.1.26] [EC:1.3.1.26]	0.00	0.00	0.00	0.00	0.95	0.35
two-component system, NtrC family, sensor histidineinase YfhK [EC:2.7.13.3] [EC:2	0.00	0.00	0.00	0.00	0.95	1.35
large subunit ribosomal protein L16 [NA]	0.00	0.00	0.00	0.00	0.93	0.00
flagellar biosynthesis protein FlhF [NA]	0.00	0.00	0.00	0.14	0.93	0.79
ubiquinol-cytochrome c reductase cytochrome b subunit [EC:1.10.2.2] [EC:1.10.2.2	0.00	0.18	0.13	0.32	0.90	0.61
flagellar M-ring protein FliF [NA]	0.00	0.14	0.10	0.00	0.90	0.62
inorganic phosphate transporter, PiT family [NA]	0.00	0.00	0.13	0.00	0.90	0.72
threonine dehydratase [EC:4.3.1.19] [EC:4.3.1.19]	0.00	0.00	0.00	0.00	0.89	1.38
adenylateinase [EC:2.7.4.3] [EC:2.7.4.3]	0.00	0.00	0.26	0.00	0.89	0.55
NADH dehydrogenase [EC:1.6.99.3] [EC:1.6.99.3]	0.00	0.18	0.00	0.16	0.89	0.54
flagellar basal-body rod protein FlgB [NA]	0.00	0.00	0.00	0.00	0.88	0.98
two-component system, OmpR family, phosphate regulon sensor histidineinase Phof	0.00	0.00	0.00	0.00	0.88	1.79
ATP-dependent RNA helicase DeaD [NA]	0.00	0.13	0.19	0.00	0.86	0.12
phosphoribosylaminoimidazole carboxylase ATPase subunit [EC:4.1.1.21] [EC:4.1.1.	0.00	0.61	0.00	0.00	0.84	0.31
flagellar basal-body rod modification protein FlgD [NA]	0.00	0.00	0.00	0.00	0.84	0.82
selenocysteine lyase [EC:4.4.1.16]	0.00	0.00	0.00	0.15	0.84	0.46
alkaline phosphatase [EC:3.1.3.1] [EC:3.1.3.1]	1.03	0.00	0.00	0.00	0.83	1.63
type IV pilus assembly protein PilE [NA]	0.00	0.00	0.00	0.00	0.82	0.15
type IV pilus assembly protein PilF [NA]	0.00	0.00	0.00	0.00	0.82	0.00
chemotaxis protein CheX [NA]	0.00	0.00	0.00	0.00	0.82	0.15
two-component system, NtrC family, nitrogen regulation response regulator GlnG [N	0.00	0.00	0.00	0.00	0.82	0.76
Na+:H+ antiporter, NhaC family [NA]	0.00	0.00	0.00	0.00	0.79	0.24
sulfate adenylyltransferase subunit 1 [EC:2.7.7.4] [EC:2.7.7.4]	0.49	0.00	0.00	0.00	0.78	1.01
phosphatidylglycerophosphatase B [EC:3.1.3.27] [EC:3.1.3.27]	0.00	0.00	0.00	0.00	0.77	0.14
chemotaxis protein CheZ [NA]	0.31	0.00	0.22	0.27	0.76	0.75
adenylylsulfateinase [EC:2.7.1.25]	0.00	0.23	0.00	0.00	0.76	0.14
2-isopropylmalate synthase [EC:2.3.3.13] [EC:2.3.3.13]	0.00	0.00	0.00	0.26	0.74	0.41
UDP-N-acetylmuramoylalanyl-D-glutamate--2, 6-diaminopimelate ligase [EC:6.3.2.1	0.15	0.15	0.00	0.00	0.74	0.68
phosphoribosylamine--glycine ligase [EC:6.3.4.13] [EC:6.3.4.13]	0.00	0.00	0.00	0.00	0.74	0.00
iron(III) transport system substrate-binding protein [NA]	0.00	0.00	0.00	0.00	0.74	0.54
arginine N-succinyltransferase [EC:2.3.1.109] [EC:2.3.1.109]	0.00	0.00	0.00	0.00	0.74	0.07
adenylosuccinate synthase [EC:6.3.4.4] [EC:6.3.4.4]	0.00	0.00	0.00	0.00	0.74	0.71
cytochrome bd-I oxidase subunit II [EC:1.10.3.-] [EC:1.10.3.-]	0.00	0.00	0.00	0.00	0.73	0.14
cytochrome c-type biogenesis protein CcmH [NA]	0.00	0.00	0.00	0.00	0.72	0.00
protein-glutamate methylesterase, two-component system, chemotaxis family, resp	0.00	0.00	0.00	0.00	0.72	1.52
RNA polymerase sigma-70 factor, ECF subfamily [NA]	0.00	0.00	0.00	0.00	0.71	1.45
aminomethyltransferase [EC:2.1.2.10] [EC:2.1.2.10]	0.00	0.00	0.00	0.00	0.71	0.06
signal peptidase II [EC:3.4.23.36] [EC:3.4.23.36]	0.00	0.00	0.00	0.37	0.70	0.64
succinate dehydrogenase cytochrome b-556 subunit [EC:1.3.99.1] [EC:1.3.99.1]	0.00	0.00	0.00	0.74	0.70	1.02
rod shape-determining protein MreC [NA]	0.00	0.00	0.00	0.25	0.69	0.17
riboflavin synthase beta chain [EC:2.5.1.-] [EC:2.5.1.-]	0.00	0.00	0.00	0.00	0.67	0.00
uroporphyrin-III C-methyltransferase [EC:2.1.1.107]	0.00	0.00	0.11	0.00	0.66	0.15
thioredoxin-like protein [NA]	0.00	0.00	0.00	0.00	0.66	0.24
two-component system, chemotaxis family, sensorinase CheA [EC:2.7.13.3] [EC:2.7	0.00	0.00	0.14	0.26	0.65	1.26
serine protease Do [EC:3.4.21.107] [EC:3.4.21.107]	0.00	0.00	0.00	0.00	0.65	0.23
1-deoxy-D-xylulose-5-phosphate reductoisomerase [EC:1.1.1.267] [EC:1.1.1.267]	0.00	0.00	0.00	0.00	0.64	0.00
ribosomal large subunit pseudouridine synthase B [EC:5.4.99.12] [EC:5.4.99.12]	0.00	0.00	0.37	0.00	0.64	0.32
3-isopropylmalate/(R)-2-methylmalate dehydratase small subunit [EC:4.2.1.33] [EC	0.00	0.00	0.00	0.00	0.63	0.82
concentrative nucleoside transporter, CNT family [NA]	0.00	0.00	0.00	0.00	0.63	0.00
general secretion pathway protein F [NA]	0.00	0.00	0.00	0.00	0.62	0.29
8-amino-7-oxononanoate synthase [EC:2.3.1.47] [EC:2.3.1.47]	0.25	0.00	0.00	0.00	0.61	0.07
2-oxoglutarate dehydrogenase E1 component [EC:1.2.4.2] [EC:1.2.4.2]	0.17	0.00	0.06	0.00	0.61	0.22
stringent starvation protein A [NA]	0.00	0.37	1.05	0.64	0.60	0.56
glucose inhibited division protein A [NA]	0.12	0.00	0.18	0.11	0.60	0.11
glucose inhibited division protein B [EC:2.1.-.-] [EC:2.1.-.-]	0.00	0.00	0.00	0.96	0.60	0.11
CaCA family Na(+)/Ca(+) antiporter	0.00	0.00	0.00	0.00	0.60	0.22
pyridoxamine 5'-phosphate oxidase [EC:1.4.3.5] [EC:1.4.3.5]	0.00	0.00	0.00	0.00	0.60	0.44
L-ascorbate oxidase [EC:1.10.3.3] [EC:1.10.3.3]	0.00	0.00	0.00	0.00	0.60	0.44
drug/metabolite transporter, DME family [NA]	0.00	0.00	0.00	0.00	0.60	0.21
citrate synthase [EC:2.3.3.1] [EC:2.3.3.1]	0.19	0.00	0.00	0.00	0.60	1.16
tRNA pseudouridine synthase B [EC:5.4.99.12] [EC:5.4.99.12]	0.00	0.00	0.00	0.00	0.59	0.07
phosphatidylserine decarboxylase [EC:4.1.1.65] [EC:4.1.1.65]	0.00	0.24	0.00	0.00	0.59	0.07
negative regulator of flagellin synthesis FlgM [NA]	0.00	0.00	0.50	0.00	0.59	2.15
type IV pilus assembly protein PilV [NA]	0.00	0.00	0.00	0.00	0.58	0.21
preprotein translocase YidC subunit [NA]	0.00	0.00	0.00	0.00	0.58	0.16
branched-chain amino acid aminotransferase [EC:2.6.1.42] [EC:2.6.1.42]	0.00	0.00	0.00	0.00	0.57	0.00
periplasmic mercuric ion binding protein [NA]	0.00	0.00	0.00	0.00	0.57	0.00
phenylalanyl-tRNA synthetase beta chain [EC:6.1.1.20] [EC:6.1.1.20]	0.10	0.19	0.07	0.17	0.56	0.47
acyl-CoA thioester hydrolase YbgC (EC:3.1.2.-)	0.00	0.00	0.00	0.59	0.55	0.00
two-component system, PleD related family, response regulator [NA]	0.00	0.00	0.07	0.17	0.55	0.26
phosphoribosylformylglycinamidine cyclo-ligase [EC:6.3.3.1] [EC:6.3.3.1]	0.00	0.00	0.00	0.00	0.55	0.00
dihydroneopterin aldolase [EC:4.1.2.25] [EC:4.1.2.25]	0.00	0.00	0.00	0.00	0.55	0.00
acetyl-CoA acyltransferase [EC:2.3.1.16] [EC:2.3.1.16]	0.00	0.19	0.00	0.17	0.54	0.83
aspartyl-tRNA synthetase [EC:6.1.1.12] [EC:6.1.1.12]	0.00	0.00	0.00	0.12	0.54	0.20
UDP-N-acetylmuramoylalanyl-D-glutamyl-2, 6-diaminopimelate--D-alanyl-D-alanine	0.17	0.16	0.00	0.14	0.54	0.30

Data represent the number of sequence hits to each target ortholog per 10,000 reads, normalized to the gene size (in base pairs) of each specific ortholog.

Table S2. Alteromonadaceae specific KEGG orthologues in control and treatment cDNAs

cytochrome bd-I oxidase subunit I [EC:1.10.3.-] [EC:1.10.3.-]	0.00	0.00	0.00	0.00	0.54	0.00
phosphoribosylaminoimidazole-succinocarboxamide synthase [EC:6.3.2.6] [EC:6.3.2	0.00	0.00	0.00	0.29	0.54	0.10
large subunit ribosomal protein L19 [NA]	0.00	0.00	0.00	0.57	0.53	0.19
signal recognition particle, subunit SRP54 [NA]	0.00	0.16	0.00	0.00	0.53	0.15
UDP-glucose 4-epimerase [EC:5.1.3.2] [EC:5.1.3.2]	0.00	0.00	0.00	0.00	0.53	0.10
UDP-N-acetylmuramoylalanine--D-glutamate ligase [EC:6.3.2.9] [EC:6.3.2.9]	0.16	0.00	0.23	0.00	0.53	0.15
chorismate synthase [EC:4.2.3.5] [EC:4.2.3.5]	0.00	0.00	0.00	0.00	0.52	0.32
flagellar P-ring protein precursor FlgI [NA]	0.00	0.00	0.00	0.00	0.52	1.96
small subunit ribosomal protein S12 [NA]	0.00	0.00	0.00	0.00	0.52	0.00
flagellar basal-body rod protein FlgF [NA]	0.00	0.00	0.00	0.55	0.51	0.19
phosphoglycerate mutase [EC:5.4.2.1] [EC:5.4.2.1]	0.32	0.00	0.00	0.00	0.51	0.00
2-oxoglutarate dehydrogenase E2 component (dihydrolipoamide succinyltransferase	0.00	0.15	0.00	0.00	0.50	0.28
small subunit ribosomal protein S1 [NA]	0.00	0.15	0.11	0.13	0.50	0.51
dimethyladenosine transferase [EC:2.1.1.-] [EC:2.1.1.-]	0.00	0.00	0.00	0.00	0.50	0.19
7,8-dihydro-8-oxoguanine triphosphatase [EC:3.6.1.-] [EC:3.6.1.-]	0.00	0.00	0.00	0.00	0.50	0.00
solute:Na+ symporter, SSS family [NA]	0.00	0.00	0.00	0.13	0.50	0.60
chemotaxis protein MotA [NA]	0.00	0.00	0.00	0.00	0.50	0.65
membrane protease subunit HflK [EC:3.4.-.-] [EC:3.4.-.-]	0.00	0.00	0.14	0.00	0.50	1.04
small subunit ribosomal protein S11 [NA]	0.00	0.00	0.00	0.00	0.49	0.00
succinyl-CoA synthetase beta subunit [EC:6.2.1.5] [EC:6.2.1.5]	0.00	0.00	0.00	0.35	0.49	0.24
HemY protein [NA]	0.00	0.00	0.00	0.35	0.49	0.12
flagellar assembly protein FliH [NA]	0.00	0.00	0.00	0.00	0.49	0.72
undecaprenyl pyrophosphate synthetase [EC:2.5.1.31] [EC:2.5.1.31]	0.00	0.00	0.00	0.00	0.49	0.09
tRNA/rRNA methyltransferase [EC:2.1.1.-] [EC:2.1.1.-]	0.00	0.00	0.21	0.00	0.49	0.36
chorismate mutase [EC:5.4.99.5]	0.00	0.00	0.00	0.00	0.49	0.48
purine-binding chemotaxis protein CheW [NA]	0.00	0.00	0.21	0.26	0.48	0.89
large subunit ribosomal protein L17 [NA]	0.00	0.00	0.00	0.00	0.48	0.17
putative ABC transport system ATP-binding protein [NA]	0.00	0.00	0.00	0.00	0.48	0.62
aminoacylase [EC:3.5.1.14] [EC:3.5.1.14]	0.00	0.00	0.00	0.00	0.48	0.18
sulfate permease, SulP family [NA]	0.20	0.19	0.00	0.17	0.47	0.64
flagellar hook-associated protein 3 FlgL [NA]	0.00	0.00	0.00	0.00	0.47	0.12
malate dehydrogenase (oxaloacetate-decarboxylating)(NADP+) [EC:1.1.1.40] [EC:1	0.00	0.00	0.00	0.00	0.46	0.97
cytochrome c-type biogenesis protein CcmF [NA]	0.00	0.00	0.00	0.00	0.46	0.24
Cu2+-exporting ATPase [EC:3.6.3.4] [EC:3.6.3.4]	0.00	0.00	0.00	0.00	0.46	0.17
flagellar protein FliS [NA]	0.00	0.00	0.00	0.00	0.46	0.17
D-3-phosphoglycerate dehydrogenase [EC:1.1.1.95] [EC:1.1.1.95]	0.00	0.00	0.13	0.00	0.46	0.56
small conductance mechanosensitive ion channel, MscS family [NA]	0.00	0.00	0.00	0.00	0.46	1.11
glycyl-tRNA synthetase beta chain [EC:6.1.1.14] [EC:6.1.1.14]	0.00	0.00	0.00	0.10	0.46	0.14
thymidylate synthase [EC:2.1.1.45] [EC:2.1.1.45]	0.00	0.00	0.00	0.00	0.46	0.25
N utilization substance protein B [NA]	0.00	0.00	0.00	0.00	0.46	1.01
enoyl-CoA hydratase [EC:4.2.1.17] [EC:4.2.1.17]	0.00	0.28	0.00	0.00	0.46	0.34
pantoate--beta-alanine ligase [EC:6.3.2.1] [EC:6.3.2.1]	0.00	0.00	0.00	0.00	0.45	0.00
large subunit ribosomal protein L1 [NA]	0.00	0.27	0.00	0.24	0.45	0.17
two-component system, chemotaxis family, response regulator CheV [NA]	0.00	0.00	0.00	0.24	0.45	0.17
histidyl-tRNA synthetase [EC:6.1.1.21] [EC:6.1.1.21]	0.37	0.00	0.13	0.00	0.45	0.27
ATP-dependent RNA helicase RhlB [EC:3.6.1.-] [EC:3.6.1.-]	0.00	0.00	0.00	0.00	0.45	0.05
F-type H+-transporting ATPase subunit gamma [EC:3.6.3.14] [EC:3.6.3.14]	0.28	0.27	1.35	0.00	0.44	0.41
riboflavininase [EC:2.7.1.26]	0.00	0.00	0.00	0.00	0.44	0.08
endonuclease [EC:3.1.30.-] [EC:3.1.30.-]	0.00	0.00	0.00	0.00	0.44	0.16
endoglucanase [EC:3.2.1.4] [EC:3.2.1.4]	0.00	0.00	0.00	0.00	0.44	1.92
alpha-glucosidase [EC:3.2.1.20] [EC:3.2.1.20]	0.00	0.00	0.00	0.00	0.44	0.16
beta-glucosidase [EC:3.2.1.21] [EC:3.2.1.21]	0.00	0.00	0.00	0.00	0.44	0.00
3-dehydroquinate dehydratase II [EC:4.2.1.10] [EC:4.2.1.10]	0.00	0.00	0.00	0.00	0.44	0.00
5-methyltetrahydrofolate--homocysteine methyltransferase [EC:2.1.1.13] [EC:2.1.1	0.00	0.00	0.06	0.08	0.44	0.27
mannose-1-phosphate guanylyltransferase [EC:2.7.7.22] [EC:2.7.7.22]	0.18	0.00	0.00	0.00	0.44	0.05
ribosomal protein L11 methyltransferase [EC:2.1.1.-] [EC:2.1.1.-]	0.00	0.00	0.00	0.00	0.43	0.08
membrane protease subunit HflC [EC:3.4.-.-] [EC:3.4.-.-]	0.00	0.00	0.00	0.00	0.43	0.16
preprotein translocase SecG subunit [NA]	0.00	0.00	0.00	0.46	0.43	0.32
hydrophobic/amphiphilic exporter-1 (mainly G- bacteria), HAE1 family [NA]	0.00	0.00	0.19	0.46	0.43	0.56
methylenetetrahydrofolate reductase (NADPH) [EC:1.5.1.20] [EC:1.5.1.20]	0.00	0.00	0.00	0.00	0.43	0.24
glutathione S-transferase [EC:2.5.1.18] [EC:2.5.1.18]	0.00	0.00	0.19	0.00	0.43	0.32
biotin carboxylase [EC:6.3.4.14]	0.00	0.00	0.00	0.00	0.43	0.00
phosphoglucosamine mutase [EC:5.4.2.10] [EC:5.4.2.10]	0.00	0.00	0.00	0.00	0.43	0.00
large subunit ribosomal protein L9 [NA]	0.00	0.00	0.00	0.00	0.43	0.63
magnesium transporter [NA]	0.00	0.00	0.00	0.00	0.43	0.10
tRNA delta(2)-isopentenylpyrophosphate transferase [EC:2.5.1.8] [EC:2.5.1.8]	0.00	0.00	0.00	0.00	0.42	1.32
DNA-directed RNA polymerase subunit beta [EC:2.7.7.6] [EC:2.7.7.6]	0.00	0.00	0.00	0.00	0.42	0.31
adenylosuccinate lyase [EC:4.3.2.2] [EC:4.3.2.2]	0.00	0.00	0.00	0.00	0.42	0.15
phosphogluconate dehydratase [EC:4.2.1.12] [EC:4.2.1.12]	0.13	0.00	0.09	0.00	0.42	0.08
regulator of sigma D [NA]	0.00	0.00	0.00	0.00	0.42	0.15
argininosuccinate lyase [EC:4.3.2.1] [EC:4.3.2.1]	0.00	0.34	0.00	0.29	0.42	0.20
flagellar hook protein FlgE [NA]	0.00	0.00	0.00	0.00	0.41	0.61
putative ABC transport system substrate-binding protein [NA]	0.00	0.00	0.00	0.43	0.41	0.60
3R-hydroxymyristoyl ACP dehydrase [EC:4.2.1.-] [EC:4.2.1.-]	0.00	0.00	0.00	0.00	0.41	0.00
indole-3-glycerol phosphate synthase [EC:4.1.1.48] [EC:4.1.1.48]	0.00	0.00	0.00	0.00	0.41	0.05
acetyl-CoA carboxylase carboxyl transferase subunit alpha [EC:6.4.1.2] [EC:6.4.1.2	0.00	0.00	0.00	0.00	0.41	0.00
general secretion pathway protein M [NA]	0.00	0.00	0.00	0.00	0.40	0.30
flagellar protein FlgJ [NA]	0.00	0.00	0.00	0.00	0.40	0.30
putative adenine-specific DNA-methyltransferase [EC:2.1.1.72] [EC:2.1.1.72]	0.00	0.00	0.00	0.00	0.40	0.22
chromosome partitioning protein [NA]	0.00	0.00	0.00	0.00	0.40	0.41
cb-type cytochrome c oxidase subunit I [EC:1.9.3.1]	0.00	0.00	0.00	0.00	0.40	0.64
M20 (carboxypeptidase Ss1) subfamily protein (EC:3.4.-.-)	0.00	0.00	0.34	0.43	0.40	1.17
DNA polymerase III subunit delta [EC:2.7.7.7] [EC:2.7.7.7]	0.00	0.00	0.00	0.00	0.40	0.00
starvation-inducible DNA-binding protein [NA]	0.00	0.00	0.00	0.00	0.39	0.29

Data represent the number of sequence hits to each target ortholog per 10,000 reads, normalized to the gene size (in base pairs) of each specific ortholog.

Table S2. Alteromonadaceae specific KEGG orthologues in control and treatment cDNAs

phosphoribosylformylglycinamidine synthase [EC:6.3.5.3] [EC:6.3.5.3]	0.00	0.00	0.00	0.05	0.39	0.07
GTP pyrophosphokinase [EC:2.7.6.5] [EC:2.7.6.5]	0.00	0.00	0.34	0.00	0.39	0.43
high-affinity choline transport protein [NA]	0.00	0.00	0.00	0.10	0.38	0.39
polysaccharide export outer membrane protein [NA]	0.47	0.00	0.33	0.00	0.38	0.14
cysteine synthase [EC:2.5.1.47]	0.24	0.00	0.00	0.00	0.38	0.28
putative amidohydrolase family protein (EC:3.5.1.-)	0.00	0.00	0.00	0.00	0.38	0.00
preprotein translocase SecB subunit [NA]	0.00	0.00	0.00	0.00	0.38	0.14
flagellin [NA]	0.47	0.00	0.00	0.00	0.38	0.83
ATP-binding cassette, subfamily B, bacterial [NA]	0.00	0.00	0.00	0.00	0.37	0.41
flagellar protein FlaG [NA]	0.00	0.00	0.00	0.00	0.37	0.41
quinolinate synthase [NA]	0.23	0.00	0.16	0.00	0.37	0.14
glutathione peroxidase [EC:1.11.1.9] [EC:1.11.1.9]	0.00	0.00	0.00	0.00	0.37	0.00
putative two-component system response regulator [NA]	0.00	0.00	0.00	0.59	0.37	0.67
disulfide bond formation protein DsbB [NA]	0.00	0.00	0.00	0.00	0.37	0.54
dihydroorotase [EC:3.5.2.3] [EC:3.5.2.3]	0.00	0.00	0.00	0.00	0.36	0.07
tetraacyldisaccharide 4'-kinase [EC:2.7.1.130] [EC:2.7.1.130]	0.00	0.00	0.00	0.00	0.36	0.34
exopolyphosphatase [EC:3.6.1.11] [EC:3.6.1.11]	0.00	0.00	0.00	0.00	0.36	0.13
flagellar motor switch protein FliM [NA]	0.00	0.00	0.00	0.00	0.36	0.13
polyribonucleotide nucleotidyltransferase [EC:2.7.7.8] [EC:2.7.7.8]	0.00	0.00	0.16	0.19	0.36	0.13
adenine phosphoribosyltransferase [EC:2.4.2.7] [EC:2.4.2.7]	0.00	0.00	0.00	0.00	0.36	0.66
hypoxanthine phosphoribosyltransferase [EC:2.4.2.8] [EC:2.4.2.8]	0.00	0.00	0.00	0.00	0.36	0.00
phosphoribosylaminoimidazolecarboxamide formyltransferase [EC:2.1.2.3]	0.00	0.00	0.00	0.00	0.36	0.00
large subunit ribosomal protein L5 [NA]	0.00	0.00	0.00	0.00	0.36	0.13
transposase [NA]	0.00	0.00	0.00	0.00	0.36	0.00
transposase [NA]	0.00	0.00	0.00	0.00	0.36	0.00
putative transposase [NA]	0.00	0.00	0.00	0.00	0.36	0.52
sucrose phosphorylase [EC:2.4.1.7] [EC:2.4.1.7]	0.00	0.00	0.00	0.00	0.35	0.13
preprotein translocase SecA subunit [NA]	0.00	0.00	0.00	0.00	0.35	0.03
transcriptional antiterminator NusG [NA]	0.00	0.42	0.00	0.00	0.35	0.38
3-isopropylmalate dehydrogenase [EC:1.1.1.85] [EC:1.1.1.85]	0.00	0.21	0.00	0.19	0.35	1.67
long-chain acyl-CoA synthetase [EC:6.2.1.3] [EC:6.2.1.3]	0.00	0.14	0.00	0.12	0.35	0.34
isoleucyl-tRNA synthetase [EC:6.1.1.5] [EC:6.1.1.5]	0.09	0.00	0.00	0.00	0.34	0.38
capsular polysaccharide transport system permease protein [NA]	0.00	0.00	0.00	0.00	0.34	0.19
hypothetical protein	0.00	0.00	0.00	0.00	0.34	0.00
ribonuclease E [EC:3.1.4.-] [EC:3.1.4.-]	0.00	0.00	0.00	0.00	0.34	0.10
D-lactate dehydrogenase [EC:1.1.1.28] [EC:1.1.1.28]	0.00	0.00	0.10	0.00	0.33	0.37
S-adenosylmethionine synthetase [EC:2.5.1.6] [EC:2.5.1.6]	0.00	0.00	0.14	0.00	0.33	0.18
membrane fusion protein [NA]	0.00	0.00	0.00	0.00	0.33	0.06
glutamate-5-semialdehyde dehydrogenase [EC:1.2.1.41] [EC:1.2.1.41]	0.00	0.00	0.00	0.00	0.33	0.61
dihydrodipicolinate synthase [EC:4.2.1.52] [EC:4.2.1.52]	0.00	0.00	0.00	0.00	0.33	0.36
peptidyl-tRNA hydrolase, PTH1 family [EC:3.1.1.29] [EC:3.1.1.29]	0.00	0.00	0.00	0.00	0.33	0.36
dihydroxy-acid dehydratase [EC:4.2.1.9] [EC:4.2.1.9]	0.14	0.00	0.00	0.00	0.33	0.49
ribonuclease D [EC:3.1.13.5] [EC:3.1.13.5]	0.00	0.00	0.00	0.00	0.33	0.18
segregation and condensation protein B [NA]	0.00	0.00	0.00	0.00	0.32	0.00
succinate dehydrogenase flavoprotein subunit [EC:1.3.99.1] [EC:1.3.99.1]	0.00	0.00	0.00	0.00	0.32	0.24
acetateinase [EC:2.7.2.1] [EC:2.7.2.1]	0.00	0.00	0.00	0.00	0.32	0.60
choline dehydrogenase [EC:1.1.99.1] [EC:1.1.99.1]	0.00	0.00	0.00	0.17	0.32	0.42
aspartate aminotransferase [EC:2.6.1.1] [EC:2.6.1.1]	0.00	0.00	0.00	0.00	0.32	0.06
guanosine-5'-triphosphate,3'-diphosphate pyrophosphatase [EC:3.6.1.40] [EC:3.6.1	0.00	0.00	0.00	0.00	0.32	0.00
ammonium transporter, Amt family [NA]	0.20	0.00	0.00	0.00	0.32	0.99
3-oxoacyl-[acyl-carrier-protein] synthase I [EC:2.3.1.41] [acyl-carrier-protein] synt	0.00	0.00	0.00	0.17	0.31	0.12
3-deoxy-7-phosphoheptulonate synthase [EC:2.5.1.54] [EC:2.5.1.54]	0.00	0.38	0.00	0.00	0.31	1.16
chemotaxis protein methyltransferase CheR [EC:2.1.1.80]	0.00	0.00	0.00	0.00	0.31	0.69
uridineinase [EC:2.7.1.48] [EC:2.7.1.48]	0.00	0.00	0.00	0.00	0.31	0.00
insulysin [EC:3.4.24.56] [EC:3.4.24.56]	0.00	0.00	0.00	0.00	0.31	0.00
ABC transporter, ATPase subunit (EC:3.6.3.25)	0.00	0.00	0.00	0.33	0.31	0.11
hypothetical protein	0.00	0.00	0.00	0.00	0.31	0.00
general secretion pathway protein D [NA]	0.00	0.00	0.00	0.00	0.31	0.41
ribonuclease T [EC:3.1.13.-] [EC:3.1.13.-]	0.00	0.00	0.00	0.00	0.30	0.00
threonine synthase [EC:4.2.3.1] [EC:4.2.3.1]	0.00	0.00	0.13	0.00	0.30	0.77
triacylglycerol lipase [EC:3.1.1.3] [EC:3.1.1.3]	0.00	0.00	0.00	0.00	0.29	0.43
alanyl-tRNA synthetase [EC:6.1.1.7] [EC:6.1.1.7]	0.00	0.00	0.00	0.08	0.29	0.51
preprotein translocase SecY subunit [NA]	0.00	0.00	0.00	0.00	0.29	0.00
cytochrome c biogenesis protein CcmG, thiol:disulfide interchange protein DsbE [NA	0.00	0.00	0.00	0.00	0.29	0.00
amino-acid N-acetyltransferase [EC:2.3.1.1] [EC:2.3.1.1]	0.00	0.00	0.00	0.00	0.29	0.22
riboflavin synthase alpha chain [EC:2.5.1.9] [EC:2.5.1.9]	0.00	0.00	0.00	0.00	0.29	0.00
ribose 5-phosphate isomerase A [EC:5.3.1.6] [EC:5.3.1.6]	0.00	0.00	0.00	0.00	0.29	0.00
glutamate-1-semialdehyde 2,1-aminomutase [EC:5.4.3.8] [EC:5.4.3.8]	0.00	0.00	0.00	0.00	0.29	0.32
peroxiredoxin (alkyl hydroperoxide reductase subunit C) [EC:1.11.1.15] [EC:1.11.1.	0.00	0.00	0.00	0.00	0.29	0.21
dTMPinase [EC:2.7.4.9] [EC:2.7.4.9]	0.00	0.00	0.00	0.00	0.29	0.00
transketolase [EC:2.2.1.1] [EC:2.2.1.1]	0.00	0.00	0.08	0.10	0.29	0.28
aconitate hydratase 1 [EC:4.2.1.3] [EC:4.2.1.3]	0.00	0.00	0.00	0.08	0.29	0.19
lipoyl(octanoyl) transferase [EC:2.3.1.181] [EC:2.3.1.181]	0.00	0.00	0.00	0.30	0.29	0.21
two-component system, OmpR family, response regulator PhoP [NA]	0.00	0.00	0.25	0.00	0.29	0.21
IMP dehydrogenase [EC:1.1.1.205] [EC:1.1.1.205]	0.00	0.00	0.00	0.00	0.28	0.05
guanosine-3',5'-bis(diphosphate) 3'-pyrophosphohydrolase [EC:3.1.7.2] [EC:3.1.7.2	0.00	0.00	0.00	0.30	0.28	0.42
bifunctional protein GlmU [EC:2.3.1.157 2.7.7.23] [EC:2.3.1.157 2.7.7.23]	0.00	0.00	0.00	0.00	0.28	0.15
3-hydroxy-3-methylglutaryl-CoA reductase [EC:1.1.1.34]	0.00	0.00	0.00	0.00	0.28	0.05
DNA repair protein RadA/Sms [NA]	0.00	0.00	0.00	0.15	0.28	0.00
6-phosphogluconolactonase [EC:3.1.1.31] [EC:3.1.1.31]	0.00	0.00	0.00	0.00	0.28	0.41
hydroxyacylglutathione hydrolase [EC:3.1.2.6] [EC:3.1.2.6]	0.00	0.00	0.00	0.00	0.28	0.00
phosphate acetyltransferase [EC:2.3.1.8]	0.00	0.11	0.00	0.00	0.28	0.10
orotidine-5'-phosphate decarboxylase [EC:4.1.1.23] [EC:4.1.1.23]	0.00	0.00	0.00	0.00	0.28	0.10
dethiobiotin synthetase [EC:6.3.3.3] [EC:6.3.3.3]	0.00	0.00	0.00	0.00	0.27	0.00

Data represent the number of sequence hits to each target ortholog per 10,000 reads, normalized to the gene size (in base pairs) of each specific ortholog.

Table S2. Alteromonadaceae specific KEGG orthologues in control and treatment cDNAs

methionyl-tRNA synthetase [EC:6.1.1.10] [EC:6.1.1.10]	0.00	0.00	0.00	0.00	0.27	0.00
DNA repair protein RecO (recombination protein O) [NA]	0.00	0.00	0.00	0.00	0.27	0.00
two-component system, OmpR family, aerobic respiration control protein ArcA [NA]	0.00	0.00	0.00	0.00	0.27	0.00
succinate dehydrogenase iron-sulfur protein [EC:1.3.99.1] [EC:1.3.99.1]	0.00	0.00	0.00	0.00	0.27	0.20
ribonuclease PH [EC:2.7.7.56] [EC:2.7.7.56]	0.00	0.00	0.00	0.29	0.27	0.10
cell division protein FtsA [NA]	0.00	0.00	0.00	0.00	0.27	0.05
amidophosphoribosyltransferase [EC:2.4.2.14] [EC:2.4.2.14]	0.00	0.00	0.00	0.14	0.27	0.20
hippurate hydrolase [EC:3.5.1.32] [EC:3.5.1.32]	0.00	0.00	0.00	0.14	0.27	0.10
UDP-N-acetylmuramate--alanine ligase [EC:6.3.2.8] [EC:6.3.2.8]	0.00	0.00	0.00	0.00	0.26	0.39
phosphoribosylformimino-5-aminoimidazole carboxamide ribotide isomerase [EC:5.3	0.32	0.00	0.00	0.00	0.26	0.29
RNA methyltransferase, TrmH family [EC:2.1.1.-] [EC:2.1.1.-]	0.00	0.00	0.00	0.00	0.26	0.09
type III secretion protein SctV [NA]	0.00	0.00	0.00	0.00	0.26	0.00
succinylglutamic semialdehyde dehydrogenase [EC:1.2.1.71] [EC:1.2.1.71]	0.00	0.00	0.00	0.00	0.26	0.19
lysine 2,3-aminomutase [EC:5.4.3.2] [EC:5.4.3.2]	0.00	0.00	0.00	0.00	0.26	0.00
N utilization substance protein A [NA]	0.00	0.00	0.00	0.14	0.26	0.23
flagellar biosynthetic protein FliP [NA]	0.00	0.00	0.00	0.00	0.25	0.28
sec-independent protein translocase protein TatC [NA]	0.00	0.31	0.00	0.00	0.25	0.09
D-beta-D-heptose 7-phosphateinase [EC:2.7.1.-]	0.00	0.00	0.00	0.00	0.25	0.00
putative copper resistance protein D [NA]	0.00	0.00	0.00	0.00	0.25	0.00
putative ABC transport system permease protein [NA]	0.00	0.00	0.21	0.00	0.25	0.45
ribosomal large subunit pseudouridine synthase A [EC:5.4.99.12] [EC:5.4.99.12]	0.00	0.00	0.00	0.00	0.25	0.00
molybdate transport system substrate-binding protein [NA]	0.00	0.00	0.00	0.00	0.25	0.09
flagellar biosynthetic protein FliR [NA]	0.00	0.00	0.21	0.00	0.25	0.18
capsular polysaccharide transport system permease protein [NA]	0.00	0.00	0.00	0.00	0.25	0.00
Na+:H+ antiporter, NhaB family [NA]	0.00	0.00	0.00	0.00	0.24	0.32
GMP synthase (glutamine-hydrolysing) [EC:6.3.5.2] [EC:6.3.5.2]	0.00	0.00	0.42	0.00	0.24	0.13
phosphatidylserine synthase [EC:2.7.8.8] [EC:2.7.8.8]	0.00	0.00	0.00	0.00	0.24	0.00
undecaprenyl-diphosphatase [EC:3.6.1.27] [EC:3.6.1.27]	0.00	0.00	0.00	0.00	0.24	0.18
methionyl-tRNA formyltransferase [EC:2.1.2.9] [EC:2.1.2.9]	0.00	0.00	0.00	0.00	0.24	0.09
penicillin amidase [EC:3.5.1.11] [EC:3.5.1.11]	0.00	0.00	0.00	0.00	0.24	0.00
Mg-dependent DNase [EC:3.1.21.-] [EC:3.1.21.-]	0.00	0.00	0.42	0.00	0.24	0.53
diaminopimelate epimerase [EC:5.1.1.7] [EC:5.1.1.7]	0.00	0.00	0.00	0.00	0.24	0.09
formamidopyrimidine-DNA glycosylase [EC:3.2.2.23 4.2.99.18] [EC:3.2.2.23 4.2.99	0.00	0.00	0.00	0.00	0.24	0.00
cyclase HisF [EC:4.1.3.-] [EC:4.1.3.-]	0.00	0.00	0.00	0.00	0.24	0.09
serine O-acetyltransferase [EC:2.3.1.30] [EC:2.3.1.30]	0.00	0.00	0.00	0.00	0.24	0.78
S-adenosylhomocysteine nucleosidase [EC:3.2.2.9]	0.00	0.00	0.00	0.00	0.24	0.08
pyrroline-5-carboxylate reductase [EC:1.5.1.2] [EC:1.5.1.2]	0.00	0.00	0.00	0.00	0.23	0.08
2,3,4,5-tetrahydropyridine-2-carboxylate N-succinyltransferase [EC:2.3.1.117] [EC:	0.00	0.00	0.00	0.00	0.23	0.51
phosphonate transport system substrate-binding protein [NA]	0.00	0.57	0.00	0.00	0.23	0.17
glycerol-3-phosphate O-acyltransferase [EC:2.3.1.15] [EC:2.3.1.15]	0.00	0.00	0.07	0.00	0.23	0.09
bis(5'-nucleosyl)-tetraphosphatase (symmetrical) [EC:3.6.1.41] [EC:3.6.1.41]	0.00	0.00	0.00	0.00	0.23	0.00
unclassified	0.00	0.00	0.20	0.00	0.23	0.17
RNA polymerase sigma-32 factor [NA]	0.00	0.00	0.00	0.00	0.23	0.42
NAD+ diphosphatase [EC:3.6.1.22] [EC:3.6.1.22]	0.00	0.00	0.00	0.00	0.23	0.00
glycerate dehydrogenase [EC:1.1.1.29] [EC:1.1.1.29]	0.00	0.00	0.00	0.00	0.23	0.00
DNA-directed RNA polymerase subunit beta' [EC:2.7.7.6] [EC:2.7.7.6]	0.06	0.00	0.08	0.05	0.23	0.15
4-amino-4-deoxychorismate lyase [EC:4.1.3.38] [EC:4.1.3.38]	0.00	0.00	0.00	0.00	0.23	0.17
gamma-glutamyltranspeptidase [EC:2.3.2.2] [EC:2.3.2.2]	0.00	0.00	0.00	0.24	0.23	0.84
saccharopine dehydrogenase (NAD+, L-glutamate forming) [EC:1.5.1.9] [EC:1.5.1.9	0.00	0.00	0.00	0.00	0.23	0.00
peptidyl-prolyl cis-trans isomerase D [EC:5.2.1.8] [EC:5.2.1.8]	0.00	0.00	0.00	0.00	0.22	0.17
flagellar biosynthesis protein FlhG [NA]	0.00	0.00	0.58	0.00	0.22	1.22
prolyl-tRNA synthetase [EC:6.1.1.15] [EC:6.1.1.15]	0.00	0.00	0.10	0.00	0.22	0.04
aspartoacylase [EC:3.5.1.15] [EC:3.5.1.15]	0.00	0.00	0.00	0.00	0.22	0.16
colicin import membrane protein [NA]	0.00	0.26	0.00	0.00	0.22	0.08
protein phosphatase 3, regulatory subunit [NA]	0.00	0.00	0.00	0.00	0.22	0.00
fructose-bisphosphate aldolase, class I [EC:4.1.2.13] [EC:4.1.2.13]	0.00	0.00	0.00	0.00	0.21	0.31
signal peptidase I [EC:3.4.21.89] [EC:3.4.21.89]	0.00	0.00	0.00	0.45	0.21	0.23
DNA mismatch repair protein MutL [NA]	0.00	0.00	0.00	0.00	0.21	0.00
general secretion pathway protein C [NA]	0.00	0.00	0.00	0.00	0.21	0.07
integrase/recombinase XerD [NA]	0.00	0.00	0.00	0.00	0.21	0.00
vitamin B12 transport system substrate-binding protein [NA]	0.00	0.00	0.00	0.00	0.21	0.00
chemotaxis protein MotB [NA]	0.00	0.00	0.00	0.00	0.21	0.38
antibiotic transport system ATP-binding protein [NA]	0.00	0.00	0.00	0.00	0.21	0.07
homoserine O-succinyltransferase [EC:2.3.1.46] [EC:2.3.1.46]	0.00	0.00	0.00	0.00	0.21	0.15
preprotein translocase SecF subunit [NA]	0.00	0.00	0.00	0.00	0.20	0.22
ribosomal large subunit pseudouridine synthase C [EC:5.4.99.12] [EC:5.4.99.12]	0.00	0.00	0.00	0.00	0.20	0.00
transaldolase [EC:2.2.1.2] [EC:2.2.1.2]	0.00	0.24	0.00	0.00	0.20	0.29
urease accessory protein [NA]	0.00	0.00	0.00	0.00	0.20	0.00
lipoic acid synthetase [EC:2.8.1.8] [EC:2.8.1.8]	0.00	0.00	0.00	0.00	0.20	0.07
arginine decarboxylase [EC:4.1.1.19] [EC:4.1.1.19]	0.00	0.12	0.00	0.00	0.20	0.22
alanine dehydrogenase [EC:1.4.1.1] [EC:1.4.1.1]	0.00	0.00	0.00	0.00	0.20	0.07
glutamate dehydrogenase [EC:1.4.1.2] [EC:1.4.1.2]	0.00	0.00	0.00	0.00	0.20	0.15
glutamate decarboxylase, putative	0.00	0.00	0.00	0.00	0.20	0.14
fructose-1,6-bisphosphatase I [EC:3.1.3.11] [EC:3.1.3.11]	0.24	0.00	0.00	0.21	0.20	0.07
glycine dehydrogenase subunit 1 [EC:1.4.4.2]	0.00	0.00	0.00	0.07	0.20	0.05
5-amino-6-(5-phosphoribosylamino)uracil reductase [EC:1.1.1.193]	0.00	0.00	0.00	0.00	0.20	0.43
phenylalanyl-tRNA synthetase alpha chain [EC:6.1.1.20] [EC:6.1.1.20]	0.00	0.00	0.00	0.00	0.20	0.14
vitamin B12 transport system ATP-binding protein [EC:3.6.3.33] [EC:3.6.3.33]	0.00	0.00	0.00	0.00	0.20	0.00
ribokinase [EC:2.7.1.15] [EC:2.7.1.15]	0.00	0.00	0.00	0.00	0.20	0.43
ribosomal large subunit pseudouridine synthase D [EC:5.4.99.12] [EC:5.4.99.12]	0.00	0.00	0.00	0.00	0.20	0.14
DNA replication and repair protein RecF [NA]	0.00	0.00	0.00	0.00	0.19	0.07
ATP-dependent DNA helicase Rep [EC:3.6.1.-] [EC:3.6.1.-]	0.00	0.00	0.00	0.00	0.19	0.00
Cu(I)/Ag(I) efflux system membrane protein CusB [NA]	0.00	0.00	0.00	0.10	0.19	0.00
glycerol-3-phosphate dehydrogenase (NAD(P)+) [EC:1.1.1.94] [EC:1.1.1.94]	0.00	0.00	0.00	0.00	0.19	0.21

Data represent the number of sequence hits to each target ortholog per 10,000 reads, normalized to the gene size (in base pairs) of each specific ortholog.

Table S2. Alteromonadaceae specific KEGG orthologues in control and treatment cDNAs

monovalent cation:H+ antiporter-2, CPA2 family [NA]	0.00	0.00	0.00	0.00	0.19	0.00
pyruvate dehydrogenase E2 component (dihydrolipoamide acetyltransferase) [EC:2.	0.00	0.00	0.00	0.00	0.19	0.07
5-formyltetrahydrofolate cyclo-ligase [EC:6.3.3.2] [EC:6.3.3.2]	0.00	0.00	0.00	0.00	0.18	0.07
flagellar motor switch protein FliG [NA]	0.00	0.00	0.00	0.00	0.18	0.34
rod shape-determining protein MreB and related proteins [NA]	0.00	0.00	0.00	0.00	0.18	0.00
succinylglutamate desuccinylase [EC:3.5.1.96] [EC:3.5.1.96]	0.00	0.00	0.00	0.00	0.18	0.00
polyphosphateinase [EC:2.7.4.1] [EC:2.7.4.1]	0.00	0.00	0.00	0.00	0.18	0.14
putative spermidine/putrescine transport system ATP-binding protein [NA]	0.00	0.00	0.00	0.00	0.18	0.00
glutamate synthase (NADPH) [EC:1.4.1.13] [EC:1.4.1.13]	0.00	0.00	0.00	0.00	0.18	0.00
aspartate carbamoyltransferase catalytic subunit [EC:2.1.3.2] [EC:2.1.3.2]	0.00	0.00	0.00	0.00	0.18	0.00
alanine racemase [EC:5.1.1.1] [EC:5.1.1.1]	0.00	0.00	0.00	0.00	0.18	0.00
ferredoxin hydrogenase [EC:1.12.7.2] [EC:1.12.7.2]	0.00	0.00	0.00	0.00	0.18	0.00
heptosyltransferase II [EC:2.4.-.-] [EC:2.4.-.-]	0.00	0.00	0.00	0.00	0.18	0.00
basic amino acid/polyamine antiporter, APA family [NA]	0.00	0.00	0.00	0.00	0.18	0.00
ribosomal RNA small subunit methyltransferase C [EC:2.1.1.52] [EC:2.1.1.52]	0.00	0.00	0.00	0.00	0.18	0.33
phospho-N-acetylmuramoyl-pentapeptide-transferase [EC:2.7.8.13] [EC:2.7.8.13]	0.00	0.00	0.00	0.00	0.18	0.06
two-component system, NtrC family, nitrogen regulation sensor histidineinase GlnL	0.00	0.00	0.00	0.00	0.18	0.39
glutamate 5-kinase [EC:2.7.2.11] [EC:2.7.2.11]	0.22	0.00	0.00	0.00	0.17	0.32
exodeoxyribonuclease V alpha subunit [EC:3.1.11.5] [EC:3.1.11.5]	0.00	0.00	0.00	0.00	0.17	0.10
tRNA (5-methylaminomethyl-2-thiouridylate)-methyltransferase [EC:2.1.1.61] [EC::	0.00	0.00	0.00	0.00	0.17	0.38
flagellar biosynthetic protein FlhB [NA]	0.00	0.00	0.00	0.00	0.17	0.00
stearoyl-CoA desaturase (delta-9 desaturase) [EC:1.14.19.1] [EC:1.14.19.1]	0.21	0.00	0.00	0.00	0.17	0.19
linoleoyl-CoA desaturase [EC:1.14.19.3] [EC:1.14.19.3]	0.21	0.00	0.00	0.00	0.17	0.00
iron complex transport system permease protein [NA]	0.00	0.00	0.00	0.00	0.17	0.06
chorismate mutase [EC:5.4.99.5]	0.00	0.00	0.00	0.00	0.17	1.53
phosphotransferase system, enzyme I, PtsP [EC:2.7.3.9] [EC:2.7.3.9]	0.00	0.00	0.00	0.00	0.16	0.12
phosphoglycerateinase [EC:2.7.2.3]	0.20	0.00	0.00	0.00	0.16	0.24
tryptophan synthase beta chain [EC:4.2.1.20] [EC:4.2.1.20]	0.00	0.00	0.00	0.00	0.16	0.06
putative dehydrogenase (EC:1.1.1.-)	0.00	0.00	0.00	0.00	0.16	0.12
[protein-PII] uridylyltransferase [EC:2.7.7.59] [protein-PII] uridylyltransferase (EC:	0.00	0.00	0.00	0.00	0.16	0.42
benzoate membrane transport protein [NA]	0.00	0.00	0.00	0.00	0.16	0.00
general secretion pathway protein L [NA]	0.00	0.00	0.00	0.00	0.16	0.00
argininosuccinate synthase [EC:6.3.4.5] [EC:6.3.4.5]	0.00	0.00	0.00	0.00	0.16	0.17
acetylornithine/N-succinyldiaminopimelate aminotransferase [EC:2.6.1.11 2.6.1.17]	0.39	0.00	0.28	0.00	0.16	0.29
erythronate-4-phosphate dehydrogenase [EC:1.1.1.290] [EC:1.1.1.290]	0.00	0.00	0.00	0.00	0.16	0.06
acetyl-CoA C-acetyltransferase [EC:2.3.1.9] [EC:2.3.1.9]	0.00	0.00	0.00	0.17	0.16	0.00
molybdopterin biosynthesis protein MoeA [NA]	0.00	0.00	0.00	0.00	0.16	0.06
3-oxoacyl-[acyl-carrier protein] reductase [EC:1.1.1.100] [acyl-carrier protein] redu	0.00	0.19	0.27	0.00	0.15	0.17
threonine 3-dehydrogenase [EC:1.1.1.103] [EC:1.1.1.103]	0.00	0.00	0.00	0.00	0.15	0.00
diaminopimelate decarboxylase [EC:4.1.1.20] [EC:4.1.1.20]	0.00	0.00	0.00	0.16	0.15	0.28
DNA segregation ATPase FtsK/SpoIIIE, S-DNA-T family [NA]	0.00	0.00	0.00	0.00	0.15	0.03
Cu(I)/Ag(I) efflux system membrane protein CusA [NA]	0.00	0.00	0.00	0.00	0.15	0.00
seryl-tRNA synthetase [EC:6.1.1.11] [EC:6.1.1.11]	0.00	0.00	0.00	0.16	0.15	0.11
glucose-1-phosphate adenylyltransferase [EC:2.7.7.27] [EC:2.7.7.27]	0.00	0.00	0.00	0.00	0.15	0.11
histidinol dehydrogenase [EC:1.1.1.23] [EC:1.1.1.23]	0.00	0.00	0.00	0.00	0.15	0.11
short-chain fatty acids transporter [NA]	0.00	0.00	0.00	0.00	0.15	0.00
ATP-dependent helicase Lhr and Lhr-like helicase [EC:3.6.1.-] [EC:3.6.1.-]	0.00	0.00	0.00	0.00	0.15	0.11
ribosomal RNA small subunit methyltransferase B [EC:2.1.1.-] [EC:2.1.1.-]	0.00	0.00	0.00	0.00	0.15	0.00
carbamoyl-phosphate synthase small subunit [EC:6.3.5.5] [EC:6.3.5.5]	0.18	0.00	0.00	0.15	0.15	0.05
succinylarginine dihydrolase [EC:3.5.3.23] [EC:3.5.3.23]	0.00	0.00	0.12	0.00	0.14	0.00
2-octaprenyl-6-methoxyphenol hydroxylase [EC:1.14.13.-] [EC:1.14.13.-]	0.00	0.00	0.00	0.00	0.14	0.00
pyruvate dehydrogenase E1 component [EC:1.2.4.1] [EC:1.2.4.1]	0.00	0.00	0.00	0.08	0.14	0.00
neurotransmitter:Na+ symporter, NSS family [NA]	0.00	0.00	0.00	0.00	0.14	0.31
cystathionine beta-lyase [EC:4.4.1.8] [EC:4.4.1.8]	0.00	0.00	0.00	0.00	0.14	0.00
cysteine desulfurase [EC:2.8.1.7] [EC:2.8.1.7]	0.35	0.00	0.00	0.15	0.14	0.21
DNA polymerase I [EC:2.7.7.7] [EC:2.7.7.7]	0.00	0.00	0.00	0.00	0.14	0.02
DNA polymerase III subunit gamma/tau [EC:2.7.7.7] [EC:2.7.7.7]	0.00	0.00	0.00	0.07	0.14	0.10
cell division protein FtsW [NA]	0.00	0.00	0.00	0.00	0.14	0.20
flagellar hook-associated protein 2 [NA]	0.17	0.00	0.00	0.14	0.14	0.65
dihydrolipoamide dehydrogenase [EC:1.8.1.4] [EC:1.8.1.4]	0.00	0.00	0.00	0.00	0.14	0.05
cell division protease FtsH [EC:3.4.24.-] [EC:3.4.24.-]	0.00	0.00	0.00	0.00	0.13	0.30
deoxyribodipyrimidine photo-lyase [EC:4.1.99.3] [EC:4.1.99.3]	0.17	0.00	0.00	0.00	0.13	0.00
chitin deacetylase [EC:3.5.1.41] [EC:3.5.1.41]	0.00	0.00	0.00	0.00	0.13	0.00
dihydrofolate synthase [EC:6.3.2.12]	0.00	0.00	0.00	0.28	0.13	0.10
trk system potassium uptake protein TrkH [NA]	0.00	0.00	0.00	0.00	0.13	0.00
glucose-6-phosphate 1-dehydrogenase [EC:1.1.1.49] [EC:1.1.1.49]	0.00	0.15	0.00	0.00	0.13	0.00
alanine or glycine:cation symporter, AGCS family [NA]	0.00	0.00	0.11	0.00	0.13	0.14
purine nucleosidase [EC:3.2.2.1] [EC:3.2.2.1]	0.00	0.00	0.11	0.00	0.13	0.09
iron(III) transport system permease protein [NA]	0.00	0.00	0.00	0.00	0.13	0.05
fumarate hydratase, class I [EC:4.2.1.2] [EC:4.2.1.2]	0.00	0.00	0.00	0.27	0.13	0.05
glycerol-3-phosphate dehydrogenase [EC:1.1.99.5] [EC:1.1.99.5]	0.00	0.00	0.00	0.00	0.13	0.00
quinoprotein glucose dehydrogenase [EC:1.1.5.2] [EC:1.1.5.2]	0.00	0.00	0.32	0.00	0.13	0.04
phosphate transport system permease protein [NA]	0.00	0.00	0.00	0.00	0.12	0.04
lysyl-tRNA synthetase, class II [EC:6.1.1.6] [EC:6.1.1.6]	0.00	0.00	0.00	0.00	0.12	0.40
two-component system, NarL family, sensor histidineinase BarA [EC:2.7.13.3] [EC::	0.08	0.00	0.00	0.00	0.12	0.25
thiamine-phosphate pyrophosphorylase [EC:2.5.1.3] [EC:2.5.1.3]	0.00	0.00	0.00	0.00	0.12	0.00
carbamoyl-phosphate synthase large subunit [EC:6.3.5.5] [EC:6.3.5.5]	0.00	0.00	0.00	0.06	0.12	0.17
CTP synthase [EC:6.3.4.2] [EC:6.3.4.2]	0.29	0.00	0.00	0.00	0.12	0.13
aspartyl-tRNA(Asn)/glutamyl-tRNA (Gln) amidotransferase subunit A [EC:6.3.5.6 6.:	0.00	0.00	0.00	0.00	0.12	0.13
type IV pilus assembly protein PilQ [NA]	0.00	0.00	0.00	0.00	0.12	0.04
two-component system, NtrC family, response regulator PilR [NA]	0.00	0.00	0.00	0.00	0.12	0.04
glutaminyl-tRNA synthetase [EC:6.1.1.18] [EC:6.1.1.18]	0.00	0.14	0.00	0.00	0.12	0.17
L-aspartate oxidase [EC:1.4.3.16] [EC:1.4.3.16]	0.00	0.00	0.10	0.00	0.11	0.08
malate dehydrogenase (oxaloacetate-decarboxylating) [EC:1.1.1.38] [EC:1.1.1.38]	0.00	0.00	0.00	0.00	0.11	0.00

Data represent the number of sequence hits to each target ortholog per 10,000 reads, normalized to the gene size (in base pairs) of each specific ortholog.

Table S2. Alteromonadaceae specific KEGG orthologues in control and treatment cDNAs

urease alpha subunit [EC:3.5.1.5] [EC:3.5.1.5]	0.00	0.00	0.00	0.00	0.11	0.00
cell division protein FtsI (penicillin binding protein 3) [EC:2.4.1.129] [EC:2.4.1.129]	0.00	0.00	0.00	0.00	0.11	0.04
transcription-repair coupling factor (superfamily II helicase) [EC:3.6.1.-] [EC:3.6.1.	0.00	0.00	0.00	0.00	0.11	0.04
asparagine synthase (glutamine-hydrolysing) [EC:6.3.5.4] [EC:6.3.5.4]	0.13	0.00	0.00	0.00	0.11	0.28
peptidyl-dipeptidase A [EC:3.4.15.1] [EC:3.4.15.1]	0.00	0.00	0.09	0.00	0.11	0.23
glucosamine--fructose-6-phosphate aminotransferase (isomerizing) [EC:2.6.1.16] [I	0.00	0.13	0.09	0.00	0.10	0.27
ATP-dependent DNA helicase DinG [EC:3.6.1.-] [EC:3.6.1.-]	0.00	0.00	0.00	0.00	0.10	0.07
ATP-dependent DNA helicase RecQ [EC:3.6.1.-] [EC:3.6.1.-]	0.00	0.00	0.00	0.00	0.10	0.04
3-methylcrotonyl-CoA carboxylase alpha subunit [EC:6.4.1.4] [EC:6.4.1.4]	0.00	0.00	0.00	0.00	0.10	0.00
exodeoxyribonuclease V beta subunit [EC:3.1.11.5] [EC:3.1.11.5]	0.00	0.00	0.00	0.05	0.10	0.00
flagellar hook-associated protein 1 FlgK [NA]	0.00	0.00	0.00	0.00	0.09	0.31
type IV pilus assembly protein PilY1 [NA]	0.00	0.00	0.00	0.00	0.09	0.07
ATP-dependent DNA helicase RecG [EC:3.6.1.-] [EC:3.6.1.-]	0.00	0.00	0.00	0.00	0.09	0.03
flagellar biosynthesis protein FlhA [NA]	0.00	0.00	0.00	0.10	0.09	0.27
preprotein translocase SecD subunit [NA]	0.00	0.11	0.00	0.00	0.09	0.03
DNA helicase II / ATP-dependent DNA helicase PcrA [EC:3.6.1.-] [EC:3.6.1.-]	0.00	0.00	0.00	0.09	0.09	0.03
4-alpha-glucanotransferase [EC:2.4.1.25] [EC:2.4.1.25]	0.00	0.11	0.08	0.00	0.09	0.06
ribonucleoside-diphosphate reductase alpha chain [EC:1.17.4.1] [EC:1.17.4.1]	0.00	0.00	0.00	0.00	0.08	0.06
aldehyde dehydrogenase (NAD+) [EC:1.2.1.3] [EC:1.2.1.3]	0.00	0.00	0.00	0.00	0.08	0.00
type IV pili sensor histidineinase and response regulator	0.00	0.00	0.00	0.00	0.08	0.00
phosphoenolpyruvate carboxylase [EC:4.1.1.31] [EC:4.1.1.31]	0.00	0.00	0.00	0.00	0.07	0.16
valyl-tRNA synthetase [EC:6.1.1.9] [EC:6.1.1.9]	0.00	0.00	0.12	0.00	0.07	0.10
ATP-dependent helicase HepA [EC:3.6.1.-] [EC:3.6.1.-]	0.00	0.00	0.00	0.00	0.07	0.10
heavy-metal exporter, HME family [NA]	0.00	0.00	0.00	0.00	0.06	0.09
chromate transporter [NA]	0.00	0.00	0.00	0.00	0.06	0.02
DNA polymerase III subunit alpha [EC:2.7.7.7] [EC:2.7.7.7]	0.00	0.00	0.00	0.00	0.06	0.00
1-pyrroline-5-carboxylate dehydrogenase [EC:1.5.1.12]	0.00	0.00	0.00	0.05	0.05	0.06
alcohol dehydrogenase [EC:1.1.1.1]	0.00	0.00	0.00	0.00	0.00	0.30
alcohol dehydrogenase (NADP+) [EC:1.1.1.2] [EC:1.1.1.2]	0.00	0.00	0.00	0.00	0.00	0.06
UDPglucose 6-dehydrogenase [EC:1.1.1.22] [EC:1.1.1.22]	0.00	0.00	0.00	0.00	0.00	0.06
3-hydroxyisobutyrate dehydrogenase [EC:1.1.1.31] [EC:1.1.1.31]	0.00	0.00	0.00	0.24	0.00	0.17
acetoacetyl-CoA reductase [EC:1.1.1.36] [EC:1.1.1.36]	0.00	0.00	0.00	0.00	0.00	0.04
isocitrate dehydrogenase (NAD+) [EC:1.1.1.41] [EC:1.1.1.41]	0.00	0.00	0.00	0.00	0.00	0.21
isocitrate dehydrogenase [EC:1.1.1.42] [EC:1.1.1.42]	0.00	0.00	0.00	0.00	0.00	0.09
dTDP-4-dehydrorhamnose reductase [EC:1.1.1.133] [EC:1.1.1.133]	0.00	0.00	0.00	0.00	0.00	0.08
UDP-N-acetylmuramate dehydrogenase [EC:1.1.1.158] [EC:1.1.1.158]	0.00	0.00	0.00	0.38	0.00	0.00
xanthine dehydrogenase [EC:1.17.1.4] [EC:1.17.1.4]	0.00	0.00	0.00	0.00	0.00	0.07
4-hydroxythreonine-4-phosphate dehydrogenase [EC:1.1.1.262] [EC:1.1.1.262]	0.00	0.00	0.17	0.00	0.00	0.07
formate dehydrogenase, alpha subunit [EC:1.2.1.2] [EC:1.2.1.2]	0.00	0.00	0.00	0.00	0.00	0.03
succinate-semialdehyde dehydrogenase (NADP+) [EC:1.2.1.16] [EC:1.2.1.16]	0.00	0.00	0.12	0.00	0.00	0.00
N-acetyl-gamma-glutamyl-phosphate reductase [EC:1.2.1.38] [EC:1.2.1.38]	0.00	0.00	0.00	0.00	0.00	0.07
unclassified	0.00	0.00	0.00	0.00	0.00	0.12
pyruvate dehydrogenase E1 component subunit beta [EC:1.2.4.1] [EC:1.2.4.1]	0.00	0.00	0.00	0.00	0.00	0.36
oxidoreductase (EC:1.3.1.-)	0.00	0.00	0.00	0.00	0.00	0.26
dihydroorotate oxidase [EC:1.3.3.1] [EC:1.3.3.1]	0.00	0.00	0.00	0.00	0.00	0.07
glutamate dehydrogenase (NADP+) [EC:1.4.1.4] [EC:1.4.1.4]	0.00	0.00	0.17	0.00	0.00	0.00
leucine dehydrogenase [EC:1.4.1.9] [EC:1.4.1.9]	0.00	0.22	0.00	0.00	0.00	0.07
dihydrofolate reductase [EC:1.5.1.3] [EC:1.5.1.3]	0.00	0.00	0.00	0.00	0.00	0.14
methylenetetrahydrofolate dehydrogenase (NADP+) [EC:1.5.1.5]	0.00	0.00	0.00	0.00	0.00	0.08
saccharopine dehydrogenase (NAD+, L-lysine forming) [EC:1.5.1.7] [EC:1.5.1.7]	0.00	0.00	0.00	0.00	0.00	0.17
NAD(P) transhydrogenase [EC:1.6.1.1] [EC:1.6.1.1]	0.00	0.00	0.00	0.00	0.00	0.05
nitrate reductase catalytic subunit [EC:1.7.99.4] [EC:1.7.99.4]	0.00	0.00	0.06	0.00	0.00	0.00
glutathione reductase (NADPH) [EC:1.8.1.7] [EC:1.8.1.7]	0.00	0.00	0.00	0.00	0.00	0.10
thioredoxin reductase (NADPH) [EC:1.8.1.9] [EC:1.8.1.9]	0.00	0.00	0.00	0.22	0.00	0.07
peroxiredoxin [EC:1.11.1.-] [EC:1.11.1.-]	0.00	0.00	0.00	0.00	0.00	0.13
homogentisate 1,2-dioxygenase [EC:1.13.11.5] [EC:1.13.11.5]	0.00	0.00	0.24	0.00	0.00	0.00
ribonucleoside-diphosphate reductase beta chain [EC:1.17.4.1] [EC:1.17.4.1]	0.00	0.00	0.00	0.00	0.00	0.06
tRNA (guanosine-2'-O-)-methyltransferase [EC:2.1.1.34] [EC:2.1.1.34]	0.34	0.00	0.00	0.00	0.00	0.00
tRNA (uracil-5-)-methyltransferase [EC:2.1.1.35] [EC:2.1.1.35]	0.00	0.00	0.00	0.00	0.00	0.10
3-demethylubiquinone-9 3-methyltransferase [EC:2.1.1.64] [EC:2.1.1.64]	0.00	0.00	0.00	0.00	0.00	0.14
protein-L-isoaspartate(D-aspartate) O-methyltransferase [EC:2.1.1.77] [EC:2.1.1.7	0.00	0.00	0.00	0.00	0.00	0.11
23S rRNA methyltransferase (EC:2.1.1.-)	0.00	0.00	0.00	0.00	0.00	0.29
glycine hydroxymethyltransferase [EC:2.1.2.1] [EC:2.1.2.1]	0.00	0.00	0.00	0.00	0.00	0.34
3-methyl-2-oxobutanoate hydroxymethyltransferase [EC:2.1.2.11] [EC:2.1.2.11]	0.00	0.29	0.00	0.00	0.00	0.09
ornithine carbamoyltransferase [EC:2.1.3.3] [EC:2.1.3.3]	0.00	0.00	0.18	0.00	0.00	1.31
glycine C-acetyltransferase [EC:2.3.1.29] [EC:2.3.1.29]	0.00	0.00	0.00	0.00	0.00	0.06
3-oxoacyl-[acyl-carrier-protein] synthase III [EC:2.3.1.180] [acyl-carrier-protein] s	0.00	0.00	0.15	0.00	0.00	0.19
UDP-N-acetylglucosamine acyltransferase [EC:2.3.1.129] [EC:2.3.1.129]	0.00	0.00	0.00	0.27	0.00	0.64
acetyl-CoA CoA transferase / acetyltransferase (EC:2.3.1.-)	0.00	0.00	0.00	0.00	0.00	0.18
starch phosphorylase [EC:2.4.1.1] [EC:2.4.1.1]	0.09	0.00	0.07	0.00	0.00	0.03
1,4-alpha-glucan branching enzyme [EC:2.4.1.18] [EC:2.4.1.18]	0.00	0.00	0.00	0.00	0.00	0.10
starch synthase [EC:2.4.1.21] [EC:2.4.1.21]	0.00	0.00	0.00	0.00	0.00	0.14
lipid-A-disaccharide synthase [EC:2.4.1.182] [EC:2.4.1.182]	0.00	0.00	0.00	0.00	0.00	0.10
putative teichoic acid/polysaccharide glycosyl transferase, group 1	0.00	0.00	0.00	0.00	0.00	0.03
orotate phosphoribosyltransferase [EC:2.4.2.10] [EC:2.4.2.10]	0.00	0.36	0.00	0.00	0.00	0.33
ATP phosphoribosyltransferase [EC:2.4.2.17] [EC:2.4.2.17]	0.00	0.26	0.00	0.23	0.00	0.00
anthranilate phosphoribosyltransferase [EC:2.4.2.18] [EC:2.4.2.18]	0.00	0.22	0.00	0.00	0.00	0.00
nicotinate-nucleotide pyrophosphorylase (carboxylating) [EC:2.4.2.19] [EC:2.4.2.19	0.00	0.00	0.00	0.00	0.00	0.08
queuine tRNA-ribosyltransferase [EC:2.4.2.29] [EC:2.4.2.29]	0.00	0.00	0.00	0.00	0.00	0.22
geranyltranstransferase [EC:2.5.1.10] [EC:2.5.1.10]	0.00	0.00	0.00	0.00	0.00	0.08
spermidine synthase [EC:2.5.1.16] [EC:2.5.1.16]	0.00	0.00	0.00	0.23	0.00	0.00
histidinol-phosphate aminotransferase [EC:2.6.1.9] [EC:2.6.1.9]	0.00	0.00	0.00	0.00	0.00	0.12
alanine--glyoxylate transaminase [EC:2.6.1.44]	0.00	0.00	0.00	0.00	0.00	0.49
glucokinase [EC:2.7.1.2] [EC:2.7.1.2]	0.00	0.00	0.00	0.21	0.00	0.30

Data represent the number of sequence hits to each target ortholog per 10,000 hits, normalized to the gene size (in base pairs) of each specific ortholog.

Table S2. Alteromonadaceae specific KEGG orthologues in control and treatment cDNAs

fructokinase [EC:2.7.1.4] [EC:2.7.1.4]	0.00	0.00	0.00	0.21	0.00	0.00
NAD+inase [EC:2.7.1.23] [EC:2.7.1.23]	0.00	0.00	0.00	0.00	0.00	0.08
glycerolinase [EC:2.7.1.30] [EC:2.7.1.30]	0.00	0.00	0.00	0.00	0.00	0.05
pyruvateinase [EC:2.7.1.40] [EC:2.7.1.40]	0.00	0.00	0.00	0.00	0.00	0.15
N-acetylglucosamineinase [EC:2.7.1.59] [EC:2.7.1.59]	0.00	0.00	0.00	0.00	0.00	0.11
4-diphosphocytidyl-2-C-methyl-D-erythritolinase [EC:2.7.1.148] [EC:2.7.1.148]	0.00	0.25	0.00	0.00	0.00	0.00
aspartateinase [EC:2.7.2.4] [EC:2.7.2.4]	0.18	0.00	0.25	0.15	0.00	0.05
guanylateinase [EC:2.7.4.8] [EC:2.7.4.8]	0.00	0.00	0.00	0.00	0.00	0.11
ribose-phosphate pyrophosphokinase [EC:2.7.6.1] [EC:2.7.6.1]	0.00	0.00	0.17	0.00	0.00	0.07
sulfate adenylyltransferase subunit 2 [EC:2.7.7.4] [EC:2.7.7.4]	0.00	0.00	0.00	0.00	0.00	0.08
UTP--glucose-1-phosphate uridylyltransferase [EC:2.7.7.9] [EC:2.7.7.9]	0.00	0.00	0.00	0.00	0.00	0.63
poly(A) polymerase [EC:2.7.7.19] [EC:2.7.7.19]	0.00	0.00	0.00	0.00	0.00	0.27
glucose-1-phosphate thymidylyltransferase [EC:2.7.7.24] [EC:2.7.7.24]	0.00	0.00	0.00	0.00	0.00	0.08
phosphatidate cytidylyltransferase [EC:2.7.7.41] [EC:2.7.7.41]	0.00	0.00	0.00	0.00	0.00	0.24
glutamate-ammonia-ligase adenylyltransferase [EC:2.7.7.42] [EC:2.7.7.42]	0.00	0.00	0.00	0.00	0.00	0.07
3-mercaptopyruvate sulfurtransferase [EC:2.8.1.2] [EC:2.8.1.2]	0.00	0.00	0.00	0.00	0.00	0.06
biotin synthetase [EC:2.8.1.6] [EC:2.8.1.6]	0.21	0.00	0.00	0.00	0.00	0.06
3-oxoacid CoA-transferase subunit A [EC:2.8.3.5] [EC:2.8.3.5]	0.00	0.00	0.00	0.00	0.00	0.40
esterase / lipase [EC:3.1.1.-] [EC:3.1.1.-]	0.00	0.00	0.00	0.00	0.00	0.10
acetyl-CoA hydrolase [EC:3.1.2.1] [EC:3.1.2.1]	0.00	0.00	0.00	0.00	0.00	0.20
palmitoyl-CoA hydrolase [EC:3.1.2.2] [EC:3.1.2.2]	0.00	0.00	0.00	0.00	0.00	0.10
phosphoserine phosphatase [EC:3.1.3.3] [EC:3.1.3.3]	0.00	0.00	0.00	0.30	0.00	0.30
histidinol-phosphatase [EC:3.1.3.15]	0.00	0.00	0.15	0.00	0.00	0.33
phosphoglycolate phosphatase [EC:3.1.3.18] [EC:3.1.3.18]	0.00	0.00	0.48	0.00	0.00	0.10
glycerophosphoryl diester phosphodiesterase [EC:3.1.4.46] [EC:3.1.4.46]	0.00	0.00	0.00	0.00	0.00	0.15
dGTPase [EC:3.1.5.1] [EC:3.1.5.1]	0.00	0.00	0.00	0.00	0.00	0.05
chitinase [EC:3.2.1.14] [EC:3.2.1.14]	0.00	0.00	0.00	0.00	0.00	0.48
beta-galactosidase [EC:3.2.1.23] [EC:3.2.1.23]	0.00	0.00	0.00	0.00	0.00	0.16
beta-N-acetylhexosaminidase [EC:3.2.1.52] [EC:3.2.1.52]	0.00	0.00	0.00	0.00	0.00	0.09
unclassified	0.00	0.00	0.00	0.00	0.00	0.05
DNA glycosylase [EC:3.2.2.-] [EC:3.2.2.-]	0.00	0.00	0.00	0.00	0.00	0.05
leukotriene-A4 hydrolase [EC:3.3.2.6] [EC:3.3.2.6]	0.00	0.00	0.00	0.00	0.00	0.38
proline iminopeptidase [EC:3.4.11.5] [EC:3.4.11.5]	0.00	0.09	0.00	0.00	0.00	0.05
methionyl aminopeptidase [EC:3.4.11.18] [EC:3.4.11.18]	0.00	0.00	0.00	0.00	0.00	0.18
L-asparaginase [EC:3.5.1.1] [EC:3.5.1.1]	0.00	0.00	0.00	0.00	0.00	0.07
glutaminase [EC:3.5.1.2] [EC:3.5.1.2]	0.26	0.00	0.00	0.00	0.00	0.00
amidase [EC:3.5.1.4] [EC:3.5.1.4]	0.00	0.00	0.00	0.00	0.00	0.15
beta-ureidopropionase [EC:3.5.1.6] [EC:3.5.1.6]	0.00	0.00	0.00	0.00	0.00	0.23
succinyl-diaminopimelate desuccinylase [EC:3.5.1.18] [EC:3.5.1.18]	0.00	0.00	0.00	0.00	0.00	0.37
N-acetylmuramoyl-L-alanine amidase [EC:3.5.1.28] [EC:3.5.1.28]	0.00	0.00	0.58	0.00	0.00	0.10
allophanate hydrolase [EC:3.5.1.54] [EC:3.5.1.54]	0.00	0.00	0.00	0.00	0.00	0.05
beta-lactamase [EC:3.5.2.6] [EC:3.5.2.6]	0.00	0.00	0.00	0.00	0.00	0.07
guanine deaminase [EC:3.5.4.3] [EC:3.5.4.3]	0.00	0.00	0.00	0.00	0.00	0.05
dCTP deaminase [EC:3.5.4.13] [EC:3.5.4.13]	0.00	0.00	0.00	0.00	0.00	0.12
GTP cyclohydrolase II [EC:3.5.4.25]	0.00	0.00	0.00	0.19	0.00	0.00
nitrilase [EC:3.5.5.1] [EC:3.5.5.1]	0.00	0.00	0.00	0.00	0.00	0.13
ADP-ribose pyrophosphatase [EC:3.6.1.13] [EC:3.6.1.13]	0.00	0.00	0.00	0.00	0.00	0.12
kynureninase [EC:3.7.1.3] [EC:3.7.1.3]	0.00	0.00	0.00	0.00	0.00	0.06
acylpyruvate hydrolase [EC:3.7.1.5] [EC:3.7.1.5]	0.00	0.00	0.00	0.00	0.00	0.11
2-haloacid dehalogenase [EC:3.8.1.2] [EC:3.8.1.2]	0.00	0.19	0.00	0.00	0.00	0.00
oxaloacetate decarboxylase, alpha subunit [EC:4.1.1.3] [EC:4.1.1.3]	0.00	0.00	0.00	0.00	0.00	0.04
oxaloacetate decarboxylase, gamma subunit [EC:4.1.1.3] [EC:4.1.1.3]	0.00	0.00	0.00	0.00	0.00	0.34
phosphoribosylaminoimidazole carboxylase catalytic subunit [EC:4.1.1.21] [EC:4.1.:	0.00	0.00	0.00	0.00	0.00	0.14
phosphopantothenoylcysteine decarboxylase [EC:4.1.1.36]	0.00	0.00	0.00	0.00	0.00	0.29
uroporphyrinogen decarboxylase [EC:4.1.1.37] [EC:4.1.1.37]	0.00	0.00	0.00	0.00	0.00	0.06
5-oxopent-3-ene-1,2,5-tricarboxylate decarboxylase [EC:4.1.1.68] [EC:4.1.1.68]	0.00	0.00	0.00	0.00	0.00	0.07
2-dehydro-3-deoxyphosphogluconate aldolase / 4-hydroxy-2-oxoglutarate aldolase	0.76	0.00	0.00	0.00	0.00	0.23
2-dehydro-3-deoxyphosphooctonate aldolase (KDO 8-P synthase) [EC:2.5.1.55] [EC	0.00	0.00	0.00	0.00	0.00	0.08
anthranilate synthase component I [EC:4.1.3.27] [EC:4.1.3.27]	0.00	0.00	0.00	0.00	0.00	0.04
2-methylcitrate synthase [EC:2.3.3.5] [EC:2.3.3.5]	0.00	0.00	0.00	0.00	0.00	0.33
1-deoxy-D-xylulose-5-phosphate synthase [EC:2.2.1.7] [EC:2.2.1.7]	0.00	0.00	0.00	0.11	0.00	0.08
para-aminobenzoate synthetase component I [EC:2.6.1.85] [EC:2.6.1.85]	0.00	0.00	0.00	0.00	0.00	0.10
carbonic anhydrase [EC:4.2.1.1] [EC:4.2.1.1]	0.00	0.00	0.00	0.00	0.00	0.22
galactonate dehydratase [EC:4.2.1.6] [EC:4.2.1.6]	0.00	0.00	0.00	0.00	0.00	0.08
enolase [EC:4.2.1.11] [EC:4.2.1.11]	0.18	0.00	0.00	0.00	0.00	0.00
tryptophan synthase alpha chain [EC:4.2.1.20] [EC:4.2.1.20]	0.00	0.00	0.00	0.00	0.00	0.36
uroporphyrinogen-III synthase [EC:4.2.1.75] [EC:4.2.1.75]	0.33	0.00	0.00	0.00	0.00	0.00
3-dehydroquinate synthase [EC:4.2.3.4] [EC:4.2.3.4]	0.00	0.22	0.00	0.00	0.00	0.13
O-acetylhomoserine (thiol)-lyase [EC:2.5.1.49] [EC:2.5.1.49]	0.00	0.00	0.00	0.34	0.00	0.29
hydroxymethylbilane synthase [EC:2.5.1.61] [EC:2.5.1.61]	0.00	0.00	0.00	0.00	0.00	0.15
L-serine dehydratase [EC:4.3.1.17] [EC:4.3.1.17]	0.00	0.00	0.00	0.00	0.00	0.10
lactoylglutathione lyase [EC:4.4.1.5] [EC:4.4.1.5]	0.00	0.00	0.00	0.00	0.00	0.10
adenylate cyclase [EC:4.6.1.1] [EC:4.6.1.1]	0.00	0.00	0.00	0.00	0.00	0.21
2-C-methyl-D-erythritol 2,4-cyclodiphosphate synthase [EC:4.6.1.12] [EC:4.6.1.12]	0.00	0.00	0.00	0.00	0.00	0.15
aspartate racemase [EC:5.1.1.13] [EC:5.1.1.13]	0.00	0.00	0.00	0.00	0.00	0.09
ribulose-phosphate 3-epimerase [EC:5.1.3.1] [EC:5.1.3.1]	0.00	0.32	0.00	0.00	0.00	0.10
dTDP-4-dehydrorhamnose 3,5-epimerase [EC:5.1.3.13] [EC:5.1.3.13]	0.00	0.00	0.00	0.00	0.00	0.39
glucose-6-phosphate isomerase [EC:5.3.1.9] [EC:5.3.1.9]	0.00	0.00	0.10	0.00	0.00	0.00
phosphomannomutase [EC:5.4.2.8] [EC:5.4.2.8]	0.00	0.00	0.00	0.00	0.00	0.09
tyrosyl-tRNA synthetase [EC:6.1.1.1] [EC:6.1.1.1]	0.00	0.00	0.00	0.00	0.00	0.12
tryptophanyl-tRNA synthetase [EC:6.1.1.2] [EC:6.1.1.2]	0.00	0.00	0.16	0.00	0.00	0.14
threonyl-tRNA synthetase [EC:6.1.1.3] [EC:6.1.1.3]	0.00	0.00	0.00	0.00	0.00	0.37
leucyl-tRNA synthetase [EC:6.1.1.4] [EC:6.1.1.4]	0.00	0.00	0.06	0.00	0.00	0.19
cysteinyl-tRNA synthetase [EC:6.1.1.16] [EC:6.1.1.16]	0.00	0.00	0.00	0.00	0.00	0.41

Data represent the number of sequence hits to each target ortholog per 10,000 reads, normalized to the gene size (in base pairs) of each specific ortholog.

Table S2. Alteromonadaceae specific KEGG orthologues in control and treatment cDNAs

glutamyl-tRNA synthetase [EC:6.1.1.17] [EC:6.1.1.17]	0.00	0.00	0.00	0.00	0.00	0.25
arginyl-tRNA synthetase [EC:6.1.1.19] [EC:6.1.1.19]	0.14	0.00	0.00	0.00	0.00	0.20
asparaginyl-tRNA synthetase [EC:6.1.1.22] [EC:6.1.1.22]	0.17	0.00	0.00	0.00	0.00	0.15
acetyl-CoA synthetase [EC:6.2.1.1] [EC:6.2.1.1]	0.00	0.00	0.00	0.00	0.00	0.04
propionyl-CoA synthetase [EC:6.2.1.17] [EC:6.2.1.17]	0.00	0.00	0.00	0.00	0.00	0.07
NAD+ synthase [EC:6.3.1.5] [EC:6.3.1.5]	0.00	0.00	0.12	0.00	0.00	0.30
glutathionylspermidine synthase [EC:6.3.1.8] [EC:6.3.1.8]	0.00	0.00	0.00	0.00	0.00	0.10
glutamate--cysteine ligase [EC:6.3.2.2] [EC:6.3.2.2]	0.00	0.00	0.00	0.00	0.00	0.04
glutathione synthase [EC:6.3.2.3] [EC:6.3.2.3]	0.00	0.00	0.00	0.00	0.00	0.07
D-alanine-D-alanine ligase [EC:6.3.2.4] [EC:6.3.2.4]	0.25	0.00	0.00	0.00	0.00	0.15
urea carboxylase [EC:6.3.4.6] [EC:6.3.4.6]	0.00	0.00	0.00	0.00	0.00	0.06
biotin-[acetyl-CoA-carboxylase] ligase [EC:6.3.4.15]	0.00	0.00	0.00	0.00	0.00	0.07
acetyl-CoA carboxylase carboxyl transferase subunit beta [EC:6.4.1.2] [EC:6.4.1.2]	0.00	0.00	0.00	0.00	0.00	0.25
DNA ligase (NAD+) [EC:6.5.1.2] [EC:6.5.1.2]	0.00	0.15	0.00	0.00	0.00	0.00
2'-5' RNA ligase [EC:6.5.1.-] [EC:6.5.1.-]	0.00	0.00	0.00	0.00	0.00	0.65
branched-chain amino acid transport system ATP-binding protein [NA]	0.00	0.00	0.00	0.00	0.00	0.10
branched-chain amino acid transport system substrate-binding protein [NA]	0.00	0.00	0.00	0.00	0.00	0.06
molybdate transport system permease protein [NA]	0.00	0.00	0.00	0.00	0.00	0.10
peptide/nickel transport system ATP-binding protein [NA]	0.00	0.00	0.00	0.00	0.00	0.14
phosphate transport system ATP-binding protein [EC:3.6.3.27] [EC:3.6.3.27]	0.00	0.00	0.00	0.00	0.00	0.18
ATP synthase protein I [NA]	0.00	0.00	0.00	0.50	0.00	0.00
heme exporter membrane protein CcmC [NA]	0.00	0.00	0.00	0.00	0.00	0.09
cytochrome c-type biogenesis protein CcmE [NA]	0.00	0.00	0.00	0.00	0.00	0.14
cytochrome c oxidase subunit II [EC:1.9.3.1] [EC:1.9.3.1]	0.00	0.00	0.00	0.22	0.00	0.00
protoheme IX farnesyltransferase [EC:2.5.1.-] [EC:2.5.1.-]	0.00	0.00	0.00	0.00	0.00	0.47
DNA polymerase III subunit chi [EC:2.7.7.7] [EC:2.7.7.7]	0.00	0.00	0.00	0.00	0.00	0.16
DNA polymerase III subunit delta' [EC:2.7.7.7] [EC:2.7.7.7]	0.00	0.00	0.00	0.00	0.00	0.08
DNA polymerase III subunit epsilon [EC:2.7.7.7] [EC:2.7.7.7]	0.00	0.00	0.00	0.00	0.00	0.12
flagella basal body P-ring formation protein FlgA [NA]	0.00	0.00	0.00	0.00	0.00	0.22
flagellar basal-body rod protein FlgC [NA]	0.00	0.00	0.00	0.54	0.00	0.55
flagellar basal-body rod protein FlgG [NA]	0.00	0.00	0.00	0.00	0.00	0.54
flagellar L-ring protein precursor FlgH [NA]	0.35	0.00	0.00	0.00	0.00	2.18
flagella synthesis protein FlgN [NA]	0.00	0.00	0.38	0.00	0.00	0.82
RNA polymerase sigma factor for flagellar operon FliA [NA]	0.00	0.00	0.00	0.00	0.00	0.29
flagellar hook-basal body complex protein FliE [NA]	0.00	0.63	0.00	0.00	0.00	0.39
flagellar FliL protein [NA]	0.00	0.00	0.35	0.00	0.00	0.74
flagellar motor switch protein FliN/FliY [NA]	0.00	0.00	0.00	0.00	0.00	0.17
flagellar biosynthetic protein FliQ [NA]	0.00	0.00	0.00	0.00	0.00	0.26
general secretion pathway protein A [NA]	0.00	0.00	0.00	0.00	0.00	0.09
general secretion pathway protein B [NA]	0.00	0.00	0.27	0.00	0.00	0.00
general secretion pathway protein E [NA]	0.00	0.00	0.00	0.00	0.00	0.43
general secretion pathway protein H [NA]	0.00	0.00	0.00	0.00	0.00	0.23
general secretion pathway protein I [NA]	0.00	0.00	0.00	0.00	0.00	0.89
general secretion pathway protein J [NA]	0.00	0.00	0.00	0.00	0.00	0.11
general secretion pathway protein N [NA]	0.00	0.00	0.00	0.27	0.00	0.00
two-component system, NtrC family, response regulator [NA]	0.00	0.00	0.00	0.00	0.00	0.03
two-component system, NtrC family, sensorinase [EC:2.7.13.3] [EC:2.7.13.3]	0.00	0.00	0.00	0.00	0.00	0.15
two-component system, OmpR family, response regulator [NA]	0.00	0.00	0.07	0.00	0.00	0.09
two-component system, unclassified family, response regulator [NA]	0.00	0.00	0.07	0.00	0.00	0.03
glutamyl-tRNA reductase [EC:1.2.1.70] [EC:1.2.1.70]	0.00	0.00	0.00	0.25	0.00	0.08
uroporphyrin-III C-methyltransferase [EC:2.1.1.107] [EC:2.1.1.107]	0.00	0.00	0.00	0.00	0.00	0.78
glutamine amidotransferase [EC:2.4.2.-] [EC:2.4.2.-]	0.00	0.00	0.00	0.00	0.00	0.22
lipid A biosynthesis lauroyl acyltransferase [EC:2.3.1.-] [EC:2.3.1.-]	0.00	0.00	0.00	0.24	0.00	0.67
UDP-3-O-[3-hydroxymyristoyl] N-acetylglucosamine deacetylase [EC:3.5.1.-] [3-hy	0.00	0.00	0.00	0.00	0.00	0.26
UDP-3-O-[3-hydroxymyristoyl] glucosamine N-acyltransferase [EC:2.3.1.-] [3-hydrc	0.00	0.00	0.00	0.20	0.00	0.00
UDP-N-acetylglucosamine--N-acetylmuramyl-(pentapeptide) pyrophosphoryl-undeca	0.00	0.00	0.00	0.00	0.00	0.26
Nif-specific regulatory protein [NA]	0.00	0.00	0.00	0.00	0.00	0.05
type IV pilus assembly protein PilC [NA]	0.00	0.00	0.00	0.00	0.00	0.29
leader peptidase (prepilin peptidase) / N-methyltransferase [EC:2.1.1.- 3.4.23.43] [	0.00	0.00	0.00	0.00	0.00	0.24
type IV pilus assembly protein PilM [NA]	0.00	0.00	0.00	0.00	0.00	0.06
type IV pilus assembly protein PilN [NA]	0.00	0.00	0.00	0.00	0.00	0.26
type IV pilus assembly protein PilP [NA]	0.00	0.00	0.00	0.00	0.00	0.14
type IV pilus assembly protein PilW [NA]	0.00	0.00	0.00	0.00	0.00	0.17
PTS system, nitrogen regulatory IIA component [EC:2.7.1.69] [EC:2.7.1.69]	0.00	0.00	0.00	0.00	0.00	0.16
ribosome-binding factor A [NA]	0.00	0.00	0.00	0.00	0.00	0.35
large subunit ribosomal protein L21 [NA]	0.00	0.75	0.00	0.00	0.00	0.22
large subunit ribosomal protein L31 [NA]	0.00	0.00	0.79	0.00	0.00	0.67
small subunit ribosomal protein S14 [NA]	0.78	0.00	0.00	0.00	0.00	0.00
DNA-directed RNA polymerase subunit omega [EC:2.7.7.6] [EC:2.7.7.6]	0.00	0.00	0.00	0.00	0.00	0.26
RNA polymerase primary sigma factor [NA]	0.00	0.00	0.00	0.00	0.00	0.15
RNA polymerase nonessential primary-like sigma factor [NA]	0.00	0.00	0.00	0.00	0.00	0.08
RNA polymerase sigma-54 factor [NA]	0.00	0.00	0.00	0.00	0.00	0.14
signal recognition particle receptor [NA]	0.00	0.00	0.00	0.00	0.00	0.09
thiamine biosynthesis ThiH [NA]	0.00	0.00	0.00	0.00	0.00	0.12
ubiquinone/menaquinone biosynthesis methyltransferase [EC:2.1.1.-] [EC:2.1.1.-]	0.00	0.00	0.00	0.00	0.00	0.18
2-octaprenyl-3-methyl-6-methoxy-1,4-benzoquinol hydroxylase [EC:1.14.13.-] [EC:	0.00	0.00	0.00	0.00	0.00	0.06
preprotein translocase YajC subunit [NA]	0.00	0.00	0.00	0.00	0.00	0.84
RNA methyltransferase, TrmA family [EC:2.1.1.-] [EC:2.1.1.-]	0.00	0.00	0.00	0.14	0.00	0.05
3-deoxy-D-manno-octulosonate 8-phosphate phosphatase (KDO 8-P phosphatase) [	0.43	0.00	0.00	0.00	0.00	0.00
OmpA-OmpF porin, OOP family [NA]	0.00	0.00	0.00	0.27	0.00	0.37
outer membrane factor, OMF family [NA]	0.00	0.00	0.00	0.27	0.00	0.28
amino acid transporter, AAT family [NA]	0.00	0.00	0.15	0.00	0.00	0.00
cation efflux system protein, CDF family [NA]	0.00	0.00	0.00	0.00	0.00	0.08
small multidrug resistance protein, SMR family [NA]	0.00	0.00	0.51	0.00	0.00	0.00

Data represent the number of sequence hits to each target ortholog per 10,000 reads, normalized to the gene size (in base pairs) of each specific ortholog.

Table S2. Alteromonadaceae specific KEGG orthologues in control and treatment cDNAs

lactate transporter, LctP family [NA]	0.00	0.00	0.00	0.00	0.00	0.04
proton-dependent oligopeptide transporter, POT family [NA]	0.00	0.00	0.00	0.14	0.00	0.19
dicarboxylate/amino acid:cation (Na+ or H+) symporter, DAACS family [NA]	0.00	0.17	0.00	0.00	0.00	0.00
multidrug resistance protein, MATE family [NA]	0.00	0.00	0.00	0.15	0.00	0.05
alkyl hydroperoxide reductase subunit F [EC:1.6.4.-] [EC:1.6.4.-]	0.00	0.00	0.10	0.00	0.00	0.00
chemotaxis protein CheC [NA]	0.51	0.00	0.36	0.00	0.00	1.37
chemotaxis protein CheD [EC:3.5.1.44] [EC:3.5.1.44]	0.00	0.00	0.00	0.00	0.00	0.28
nucleobase:cation symporter-1, NCS1 family [NA]	0.00	0.00	0.00	0.00	0.00	0.03
ribonuclease HII [EC:3.1.26.4] [EC:3.1.26.4]	0.00	0.00	0.00	0.36	0.00	0.00
pyridoxine 5-phosphate synthase [EC:2.6.99.2] [EC:2.6.99.2]	0.00	0.00	0.00	0.00	0.00	0.09
type III pantothenateinase [EC:2.7.1.33] [EC:2.7.1.33]	0.00	0.00	0.00	0.00	0.00	0.09
4-hydroxy-3-methylbut-2-en-1-yl diphosphate synthase [EC:1.17.4.3] [EC:1.17.4.3	0.00	0.21	0.00	0.00	0.00	0.13
4-hydroxy-3-methylbut-2-enyl diphosphate reductase [EC:1.17.1.2] [EC:1.17.1.2]	0.00	0.00	0.00	0.00	0.00	0.15
cell division protein ZipA [NA]	0.00	0.00	0.00	0.00	0.00	0.09
chromosome segregation protein [NA]	0.00	0.00	0.00	0.00	0.00	0.12
cell division protein FtsZ [NA]	0.00	0.00	0.00	0.00	0.00	0.08
exonuclease SbcC [NA]	0.00	0.13	0.00	0.00	0.00	0.00
exonuclease SbcD [NA]	0.00	0.15	0.00	0.00	0.00	0.09
putative permease [NA]	0.00	0.00	0.00	0.00	0.00	0.06
holliday junction DNA helicase RuvB [NA]	0.24	0.00	0.00	0.00	0.00	0.00
recombination associated protein RdgC [NA]	0.00	0.00	0.00	0.00	0.00	0.07
carbon storage regulator [NA]	0.00	0.00	0.00	0.00	0.00	0.37
peroxiredoxin Q/BCP [EC:1.11.1.15] [EC:1.11.1.15]	0.00	0.00	0.00	0.00	0.00	0.15
A/G-specific adenine glycosylase [EC:3.2.2.-] [EC:3.2.2.-]	0.00	0.00	0.00	0.00	0.00	0.06
ATP-dependent helicase HrpA [EC:3.6.1.-] [EC:3.6.1.-]	0.06	0.00	0.00	0.00	0.00	0.04
exodeoxyribonuclease V gamma subunit [EC:3.1.11.5] [EC:3.1.11.5]	0.00	0.00	0.00	0.00	0.00	0.06
cell division protein FtsQ [NA]	0.00	0.00	0.00	0.00	0.00	0.13
ATP-binding protein involved in chromosome partitioning [NA]	0.00	0.00	0.00	0.00	0.00	0.06
exodeoxyribonuclease VII small subunit [EC:3.1.11.6] [EC:3.1.11.6]	0.00	0.00	0.00	0.00	0.00	0.05
cell division topological specificity factor [NA]	0.00	0.00	0.00	0.00	0.00	0.27
septum site-determining protein MinC [NA]	0.00	0.00	0.00	0.00	0.00	0.18
molybdenum cofactor biosynthesis protein E [NA]	0.00	0.00	0.00	0.00	0.00	0.15
molybdenum cofactor biosynthesis protein D [NA]	0.00	0.00	0.00	0.00	0.00	0.30
molybdenum cofactor biosynthesis protein C [NA]	0.00	0.00	0.00	0.00	0.00	0.14
molybdenum cofactor biosynthesis protein B [NA]	0.00	0.00	0.00	0.39	0.00	0.13
molybdenum cofactor biosynthesis protein A [NA]	0.00	0.00	0.00	0.00	0.00	0.16
uracil-DNA glycosylase [EC:3.2.2.-] [EC:3.2.2.-]	0.00	0.00	0.00	0.00	0.00	0.32
SsrA-binding protein [NA]	0.00	0.00	0.00	0.00	0.00	0.14
excinuclease ABC subunit A [NA]	0.00	0.00	0.00	0.00	0.00	0.11
excinuclease ABC subunit B [NA]	0.00	0.00	0.00	0.10	0.00	0.00
excinuclease ABC subunit C [NA]	0.00	0.00	0.00	0.00	0.00	0.13
molybdopterin biosynthesis protein MoeB [NA]	0.00	0.00	0.00	0.27	0.00	0.00
molybdopterin-guanine dinucleotide biosynthesis protein A [NA]	0.00	0.00	0.00	0.00	0.00	0.11
FKBP-type peptidyl-prolyl cis-trans isomerase SlpA [EC:5.2.1.8] [EC:5.2.1.8]	0.00	0.00	0.00	0.00	0.00	0.15
D-lactate dehydrogenase [EC:1.1.1.28] [EC:1.1.1.28]	0.00	0.00	0.00	0.00	0.00	0.14
catalase/peroxidase [EC:1.11.1.6] [EC:1.11.1.6]	0.00	0.00	0.07	0.00	0.00	0.00
ribosomal-protein-alanine N-acetyltransferase [EC:2.3.1.128] [EC:2.3.1.128]	0.49	0.00	0.00	0.42	0.00	0.00
ribosomal-protein-alanine N-acetyltransferase [EC:2.3.1.128] [EC:2.3.1.128]	0.00	0.00	0.34	0.00	0.00	0.00
virulence factor [NA]	0.00	0.00	0.00	0.00	0.00	0.09
ethanolamine utilization protein EutA [NA]	0.00	0.32	0.00	0.00	0.00	0.00
hypothetical chaperone protein [NA]	0.00	0.00	0.00	0.00	0.00	0.05
osmotically inducible protein OsmC [NA]	0.00	0.53	0.00	0.00	0.00	0.33
cell cycle protein MesJ [EC:6.3.4.-] [EC:6.3.4.-]	0.00	0.00	0.00	0.00	0.00	0.11
molecular chaperone HtpG [NA]	0.00	0.00	0.00	0.00	0.00	0.15
thiol:disulfide interchange protein DsbD [EC:1.8.1.8] [EC:1.8.1.8]	0.00	0.00	0.00	0.00	0.00	0.11
tRNA 2-thiouridine synthesizing protein A [EC:2.8.1.-] [EC:2.8.1.-]	0.00	0.00	0.00	0.00	0.00	0.27
lysyl-tRNA synthetase, class II [EC:6.1.1.6] [EC:6.1.1.6]	0.00	0.00	0.00	0.00	0.00	0.07
nitrogen regulatory protein P-II 1 [NA]	0.00	0.00	0.00	0.00	0.00	0.84
gluconate dehydratase [EC:4.2.1.39] [EC:4.2.1.39]	0.00	0.00	0.00	0.00	0.00	0.07
penicillin binding protein 1B [EC:2.4.1.129 3.4.-.-] [EC:2.4.1.129 3.4.-.-]	0.00	0.00	0.00	0.00	0.00	0.13
tRNA-dihydrouridine synthase A [EC:1.-.-.-] [EC:1.-.-.-]	0.00	0.00	0.00	0.00	0.00	0.14
tRNA-dihydrouridine synthase C [EC:1.-.-.-] [EC:1.-.-.-]	0.00	0.00	0.00	0.22	0.00	0.00
ATP-dependent RNA helicase SrmB [EC:2.7.7.-] [EC:2.7.7.-]	0.00	0.00	0.00	0.00	0.00	0.06
ATP-independent RNA helicase DbpA [NA]	0.00	0.00	0.00	0.00	0.00	0.15
multiple antibiotic resistance protein [NA]	0.00	0.00	0.00	0.00	0.00	0.12
molybdate transport system ATP-binding protein [NA]	0.00	0.00	0.00	0.00	0.00	0.05
benzoate 1,2-dioxygenase electron transfer component [NA]	0.00	0.00	0.00	0.00	0.00	0.06
potassium efflux system proteinefA [NA]	0.00	0.00	0.00	0.00	0.00	0.08
putative sigma-54 modulation protein [NA]	0.00	0.00	0.00	0.00	0.00	0.24
ribosomal protein S6 modification protein [NA]	0.00	0.00	0.00	0.00	0.00	0.08
adenylate cyclase, class 1 [EC:4.6.1.1] [EC:4.6.1.1]	0.00	0.00	0.00	0.00	0.00	0.03
methyl-accepting chemotaxis protein I, serine sensor receptor [NA]	0.00	0.00	0.00	0.00	0.00	0.03
long-chain fatty acid transport protein [NA]	0.00	0.00	0.00	0.15	0.00	0.10
ATP-binding cassette, sub-family F, member 3 [NA]	0.00	0.00	0.00	0.00	0.00	0.47
bifunctional enzyme involved in thiolation and methylation of tRNA [NA]	0.00	0.00	0.08	0.00	0.00	0.00
tRNA pseudouridine synthase A [EC:5.4.99.12] [EC:5.4.99.12]	0.00	0.00	0.00	0.00	0.00	0.18
tRNA pseudouridine synthase D [EC:5.4.99.12] [EC:5.4.99.12]	0.00	0.00	0.00	0.00	0.00	0.32
ribosomal large subunit pseudouridine synthase F [EC:5.4.99.12] [EC:5.4.99.12]	0.00	0.00	0.00	0.00	0.00	0.19
magnesium and cobalt transporter [NA]	0.00	0.00	0.00	0.00	0.00	2.30
formate transporter [NA]	0.00	0.00	0.00	0.00	0.00	0.08
ATP-dependent Clp protease adaptor protein ClpS [NA]	0.00	0.00	0.00	0.00	0.00	0.22
zinc transporter, ZIP family [NA]	0.00	0.32	0.00	0.00	0.00	0.10
peptide-methionine (S)-S-oxide reductase [EC:1.8.4.11] [EC:1.8.4.11]	0.00	0.00	0.00	0.00	0.00	0.03
putative toluene tolerance protein [NA]	0.00	0.00	0.00	0.00	0.00	0.11

Data represent the number of sequence hits to each target ortholog per 10,000 reads, normalized to the gene size (in base pairs) of each specific ortholog.

Table S2. Alteromonadaceae specific KEGG orthologues in control and treatment cDNAs

magnesium chelatase family protein [NA]	0.49	0.00	0.00	0.00	0.00	0.14
xanthine dehydrogenase accessory factor [NA]	0.00	0.00	0.00	0.00	0.00	0.15
single-stranded-DNA-specific exonuclease [EC:3.1.-.-] [EC:3.1.-.-]	0.00	0.00	0.00	0.00	0.00	0.20
transposase [NA]	0.00	0.00	0.00	0.00	0.00	0.26
transposase [NA]	0.00	0.00	0.00	0.38	0.00	0.00
putative transposase [NA]	0.00	0.00	0.00	0.00	0.00	0.13
putative translation factor [NA]	0.00	0.00	0.00	0.00	0.00	0.23
S-adenosylmethionine:tRNA ribosyltransferase-isomerase [EC:5.-.-.-] [EC:5.-.-.-]	0.00	0.00	0.00	0.00	0.00	0.07
putative RNA-binding protein containingH domain [NA]	0.00	0.00	0.00	0.00	0.00	0.48
two-component system, OmpR family, heavy metal sensor histidineinase CusS [EC::	0.00	0.00	0.00	0.00	0.00	0.10
two-component system, OmpR family, sensor histidineinasedpD [EC:2.7.13.3] [EC::	0.00	0.00	0.00	0.00	0.00	0.06
two-component system, OmpR family, sensor histidineinase TorS [EC:2.7.13.3] [EC	0.00	0.00	0.00	0.00	0.00	0.11
two-component system, OmpR family, aerobic respiration control sensor histidineina	0.00	0.00	0.00	0.00	0.00	0.27
two-component system, OmpR family, phosphate regulon response regulator OmpR	0.00	0.00	0.00	0.00	0.00	0.20
two-component system, OmpR family, copper resistance phosphate regulon respons	0.00	0.00	0.00	0.00	0.00	0.41
two-component system, NarL family, nitrate/nitrite sensor histidineinase NarX [EC:2	0.00	0.00	0.00	0.00	0.00	0.12
two-component system, NarL family, sensor histidineinase EvgS [EC:2.7.13.3] [EC::	0.00	0.00	0.00	0.00	0.00	0.02
two-component system, NarL family, invasion response regulator UvrY [NA]	0.00	0.00	0.00	0.00	0.00	0.22
UDP-4-amino-4-deoxy-L-arabinose-oxoglutarate aminotransferase [EC:2.6.1.-] [EC:	0.00	0.00	0.00	0.00	0.00	0.07
3-hexulose-6-phosphate synthase [EC:4.1.2.-] [EC:4.1.2.-]	0.79	0.00	0.00	0.00	0.00	0.35
ribonuclease G [EC:3.1.4.-] [EC:3.1.4.-]	0.00	0.00	0.00	0.14	0.00	0.00
biotin sulfoxide reductase [EC:1.-.-.-] [EC:1.-.-.-]	0.00	0.00	0.00	0.00	0.00	0.11
5'-nucleotidase [EC:3.1.3.5] [EC:3.1.3.5]	0.00	0.00	0.00	0.00	0.00	0.12
3-oxoacyl-[acyl-carrier-protein] synthase II [EC:2.3.1.179] [acyl-carrier-protein] sy	0.00	0.00	0.00	0.00	0.00	0.35
antibiotic transport system permease protein [NA]	0.00	0.00	0.00	0.00	0.00	0.07
lipoprotein-releasing system permease protein [NA]	0.00	0.00	0.00	0.00	0.00	0.11
cell division transport system permease protein [NA]	0.00	0.00	0.00	0.00	0.00	0.07
uridylateinase [EC:2.7.4.22] [EC:2.7.4.22]	0.00	0.00	0.00	0.00	0.00	0.15
monooxygenase [EC:1.14.13.-] [EC:1.14.13.-]	0.00	0.00	0.00	0.00	0.00	0.06
myosin heavy chain [NA]	0.00	0.00	0.00	0.00	0.00	0.06
N-ethylmaleimide reductase [EC:1.-.-.-] [EC:1.-.-.-]	0.00	0.00	0.00	0.00	0.00	0.43
endonuclease III [EC:4.2.99.18] [EC:4.2.99.18]	0.00	0.00	0.00	0.00	0.00	0.10
acyl-CoA thioesterase II [EC:3.1.2.-] [EC:3.1.2.-]	0.00	0.00	0.00	0.00	0.00	0.08

Data represent the number of sequence hits to each target ortholog per 10,000 reads, normalized to the gene size (in base pairs) of each specific ortholog.

Table S3 : Poisson ANOVA of Control Functional Annotations*

KO	Functional category	Pathway	ORF Annotation	AIC	Intercept Coefficient	T0 Coefficient	*p*-value	*q*-value	T2 Coefficient	*p*-value	*q*-value	T12 Coefficient	*p*-value	*q*-value
K00616	01110 Carbohydrate Metabolism	00030 Pentose phosphate pathway	transaldolase [EC:2.2.1.2]	18.22	-32.11	24.23	7.10E-01	8.18E-01	22.66	4.11E-02	4.76E-01	**25.20**	**1.86E-06**	**4.13E-04**
K01238		00530 Aminosugars metabolism	membrane-bound lytic murein transglycosylase B [EC:3.2.1.-]	16.86	-9.11	**1.69**	**1.52E-04**	**1.87E-02**	-0.34	8.70E-01	9.79E-01	-22.66	1.42E-01	5.55E-01
K02109	01120 Energy Metabolism	00195 Photosynthesis	F-type H-transporting ATPase subunit b [EC:3.6.3.14]	29.86	-5.66	**-0.98**	**3.33E-04**	**2.54E-02**	0.06	1.69E-02	4.49E-01	**-1.51**	**1.86E-07**	**5.86E-05**
K02634			photosystem I subunit III	23.02	-8.01	**1.24**	**1.96E-04**	**1.87E-02**	0.17	5.87E-01	7.89E-01	-0.36	6.06E-01	9.91E-01
K02703			photosystem II PsbA protein	33.81	-3.94	**-1.59**	**1.69E-33**	**4.13E-30**	**0.39**	**1.09E-06**	**6.59E-04**	**-1.43**	**9.22E-33**	**9.10E-30**
K02705			photosystem II PsbC protein	33.04	-4.89	**-0.62**	**2.76E-04**	**2.25E-02**	-0.07	2.70E-01	6.23E-01	**-0.61**	**9.65E-05**	**1.68E-02**
K02706			photosystem II PsbD protein	30.35	-5.30	**-1.08**	**8.85E-06**	**1.08E-02**	-0.19	4.46E-01	6.81E-01	**-1.07**	**6.88E-07**	**1.68E-04**
K00927		00710 Carbon fixation	phosphoglycerateinase [EC:2.7.2.3]	18.81	-7.32	-0.89	6.48E-01	7.58E-01	-1.03	4.26E-01	6.55E-01	**-24.45**	**3.20E-04**	**4.11E-02**
K01601			ribulose-bisphosphate carboxylase large chain [EC:4.1.1.39]	31.88	-4.70	-0.06	3.88E-03	1.11E-01	**-0.84**	**1.62E-04**	**3.04E-02**	**-3.16**	**1.12E-32**	**9.10E-30**
K01602			ribulose-bisphosphate carboxylase small chain [EC:4.1.1.39]	23.46	-5.91	-0.96	6.89E-02	4.17E-01	-0.44	7.12E-01	8.89E-01	**-25.86**	**3.53E-13**	**1.44E-10**
K00381		00920 Sulfur metabolism	sulfite reductase (NADPH) hemoprotein beta-component [EC:1.8.1.2]	18.21	-8.01	**-24.11**	**8.00E-04**	**4.82E-02**	-0.74	1.15E-01	4.76E-01	0.62	2.46E-01	6.01E-01
K09458	01130 Lipid Metabolism	00061 Fatty acid biosynthesis	3-oxoacyl-[acyl-carrier-protein] synthase II [EC:2.3.1.179]	26.13	-8.01	1.49	1.75E-01	5.17E-01	**2.00**	**7.30E-08**	**9.02E-05**	0.74	1.55E-01	5.55E-01
K01963			acetyl-CoA carboxylase carboxyl transferase subunit beta [EC:6.4.1.2]	19.41	-8.70	**1.45**	**1.52E-04**	**1.87E-02**	-0.74	6.46E-01	8.16E-01	-0.76	4.86E-01	8.19E-01
K01632		00071 Fatty acid metabolism	enoyl-CoA hydratase [EC:4.2.1.17]	27.37	-6.81	**-0.51**	**2.22E-04**	**1.87E-02**	0.45	9.47E-01	1.00E+00	0.84	2.62E-03	1.42E-01
K00147	01150 Amino Acid Metabolism	00220 Urea cycle and metabolism of amino groups	glutamate-5-semialdehyde dehydrogenase [EC:1.2.1.41]	11.94	-33.11	-0.01	1.78E-02	2.57E-01	0.36	1.93E-02	4.49E-01	**25.72**	**1.84E-04**	**2.80E-02**
K00315		00260 Glycine, serine and threonine metabolism	dimethylglycine dehydrogenase [EC:1.5.99.2]	23.86	-8.19	0.68	6.56E-01	7.66E-01	0.36	1.73E-01	5.74E-01	**1.62**	**3.48E-04**	**4.25E-02**
K00282			glycine dehydrogenase subunit 1 [EC:1.4.4.2]	23.12	-8.19	**1.42**	**1.96E-04**	**1.87E-02**	0.36	5.87E-01	7.89E-01	0.11	8.68E-01	1.00E+00
K01740		00272 Cysteine metabolism	O-acetylhomoserine (thiol)-lyase [EC:2.5.1.49]	28.56	-6.62	**0.81**	**1.34E-04**	**1.87E-02**	0.17	6.32E-01	8.65E-01	0.15	6.15E-01	1.00E+00
K00167		00280 Valine, leucine and isoleucine degradation	2-oxoisovalerate dehydrogenase E1 component, beta subunit [EC:1.2.4.4]	11.66	-33.11	**25.08**	**1.32E-04**	**1.87E-02**	0.36	1.00E+00	1.00E+00	0.33	1.00E+00	1.00E+00
K03147	01190 Metabolism of Cofactors and Vitamins	00730 Thiamine metabolism	thiamine biosynthesis protein ThiC	13.19	-32.11	23.90	5.48E-01	7.58E-01	24.45	3.62E-01	6.29E-01	**24.83**	**7.26E-05**	**1.37E-02**
K02227		00860 Porphyrin and chlorophyll metabolism	adenosylcobinamide-phosphate synthase CobD [EC:6.3.1.10]	11.81	-33.11	**25.23**	**3.64E-05**	**1.87E-02**	0.36	1.00E+00	1.00E+00	0.33	1.00E+00	1.00E+00
K04038			light-independent protochlorophyllide reductase subunit N [EC:1.18.-.-]	13.94	-33.11	-0.01	1.20E-02	2.05E-01	23.66	1.96E-01	5.86E-01	**25.72**	**1.84E-04**	**2.80E-02**
K01045	01196 Xenobiotics Biodegradation and Metabolism	00363 Bisphenol A degradation	arylesterase [EC:3.1.1.2]	11.48	-33.11	**24.90**	**4.83E-04**	**3.37E-02**	0.36	1.00E+00	1.00E+00	0.33	1.00E+00	1.00E+00
K03040	01210 Transcription	03020 RNA polymerase	DNA-directed RNA polymerase subunit alpha [EC:2.7.7.6]	27.42	-7.03	**1.00**	**1.73E-04**	**1.87E-02**	0.42	2.88E-01	6.29E-01	0.20	5.83E-01	9.69E-01
K02863	01220 Translation	03010 Ribosome	large subunit ribosomal protein L1	27.79	-7.16	0.57	5.22E-01	6.77E-01	**1.27**	**1.54E-04**	**3.04E-02**	0.69	4.49E-02	4.98E-01
K02878			large subunit ribosomal protein L16	24.58	-8.19	0.86	4.26E-01	5.88E-01	**2.01**	**2.15E-07**	**1.75E-04**	0.52	3.92E-01	7.21E-01
K02864			large subunit ribosomal protein L19	24.73	-6.19	-1.58	1.88E-03	7.75E-02	-1.46	3.99E-03	2.95E-01	**-1.33**	**2.24E-04**	**3.21E-02**
K02933			large subunit ribosomal protein L5	24.15	-7.61	**0.97**	**5.73E-04**	**3.89E-02**	0.11	5.61E-01	7.65E-01	-0.48	4.16E-01	7.29E-01
K02935			large subunit ribosomal protein L7/L12	29.16	-6.19	0.46	2.05E-02	2.80E-01	**0.64**	**4.84E-06**	**1.64E-03**	-1.06	1.40E-03	1.32E-01
K02952			small subunit ribosomal protein S13	26.46	-6.51	0.27	9.65E-02	4.29E-01	**0.49**	**2.36E-04**	**4.12E-02**	**-2.27**	**2.11E-05**	**4.30E-03**
K02986			small subunit ribosomal protein S4	26.88	-6.76	**0.73**	**4.26E-05**	**1.87E-02**	0.02	4.02E-01	6.29E-01	-0.76	6.45E-02	4.98E-01
K02994			small subunit ribosomal protein S8	24.74	-8.01	**1.56**	**5.30E-05**	**1.87E-02**	0.64	4.56E-01	6.93E-01	0.62	2.46E-01	6.01E-01
K07568		03014 Other translation proteins	S-adenosylmethionine:tRNA ribosyltransferase-isomerase [EC:5.-.-.-]	18.14	-8.70	**-23.42**	**8.00E-04**	**4.82E-02**	1.56	5.51E-03	2.96E-01	0.33	6.82E-01	1.00E+00
K01999	01310 Membrane Transport	02010 ABC transporters	branched-chain amino acid transport system substrate-binding protein	29.27	-5.43	-0.72	1.70E-01	5.06E-01	**-1.45**	**1.92E-06**	**9.41E-04**	-0.74	4.80E-04	5.59E-02
K09969			general L-amino acid transport system substrate-binding protein	28.74	-5.71	-0.16	6.30E-02	3.84E-01	-1.25	5.21E-04	8.49E-02	**-0.93**	**2.00E-04**	**3.89E-02**
K02035			peptide/nickel transport system substrate-binding protein	28.62	-5.89	-0.46	2.71E-01	5.57E-01	**-1.25**	**1.29E-05**	**3.95E-03**	-0.02	9.21E-01	1.00E+00
K03320		02052 Other ion-coupled transporters	ammonium transporter, Amt family	36.55	-4.60	**-0.87**	**7.48E-04**	**4.81E-02**	**-0.72**	**3.32E-23**	**8.13E-20**	**0.84**	**5.62E-20**	**3.43E-17**
K03307			solute:Na symporter, SSS family	31.84	-4.88	-0.44	7.57E-01	8.58E-01	**-1.11**	**6.14E-06**	**2.14E-03**	**-1.01**	**4.26E-09**	**1.49E-06**
K07648	01320 Signal Transduction	02020 Two-component system	two-component system, OmpR family, aerobic respiration control sensor histidineinase ArcB [EC:2.7.13.3]	13.81	-9.80	**1.93**	**6.17E-04**	**4.87E-02**	-22.95	4.08E-01	6.29E-01	-22.97	2.99E-01	6.01E-01
K06640	01420 Cell Growth and Death	04110 Cell cycle	ataxia telangiectasia and Rad3 related [EC:2.7.11.1]	11.48	-33.11	**24.90**	**4.83E-04**	**3.37E-02**	0.36	1.00E+00	1.00E+00	0.33	1.00E+00	1.00E+00
K01367		04210 Apoptosis	calpain-1 [EC:3.4.22.52]	11.66	-33.11	**25.08**	**1.32E-04**	**1.87E-02**	0.36	1.00E+00	1.00E+00	0.33	1.00E+00	1.00E+00
K03495		04410 Cell division	glucose inhibited division protein A	19.68	-7.72	**-24.40**	**5.90E-05**	**1.87E-02**	0.36	5.47E-01	7.47E-01	0.20	6.98E-01	1.00E+00
K04077	01530 Metabolic Disorders	04940 Type I diabetes mellitus	chaperonin GroEL	34.08	-5.15	0.02	3.56E-02	3.23E-01	**0.57**	**2.52E-05**	**6.40E-03**	0.19	1.82E-01	5.55E-01

*KO = KEGG ortholog number; AIC = Akaike Information Criterion; *q*-value is a calibration of the table-wide false discovery rate (Storey et al., 2003); See methods for definitions of the coefficients.

Table S4: Pairwise tests of Functional Annotations Between Controls (Pooled) and HMWDOM Treatments 2 Hours Post Addition *

KO	Functional category	Pathway	ORF Annotation	Control	DOM	ln(Fold change)	p-value	q-value
K01676	01110 Carbohydrate Metabolism	00020 Citrate cycle (TCA cycle)	fumarate hydratase, class I [EC:4.2.1.2]	4	9	-3.24	1.20E-04	4.92E-03
K00030			isocitrate dehydrogenase (NAD) [EC:1.1.1.41]	6	10	-2.81	1.75E-04	6.33E-03
K01681			aconitate hydratase 1 [EC:4.2.1.3]	29	32	-2.21	6.72E-09	8.44E-07
K01007		00620 Pyruvate metabolism	pyruvate,water dikinase [EC:2.7.9.2]	2	15	-4.98	1.65E-09	3.35E-07
K01571			oxaloacetate decarboxylase, alpha subunit [EC:4.1.1.3]	5	19	-4.00	3.82E-10	9.15E-08
K01572			oxaloacetate decarboxylase, beta subunit [EC:4.1.1.3]	6	11	-2.95	5.15E-05	2.66E-03
K01638			malate synthase [EC:2.3.3.9]	27	26	-2.02	9.49E-07	6.95E-05
K01637		00630 Glyoxylate and dicarboxylate metabolism	isocitrate lyase [EC:4.1.3.1]	62	40	-1.44	3.20E-06	2.28E-04
K00404	01120 Energy Metabolism	00190 Oxidative phosphorylation	cb-type cytochrome c oxidase subunit I [EC:1.9.3.1]	0	5	NA	2.63E-04	8.99E-03
K02690		00195 Photosynthesis	photosystem I core protein Ib	627	96	0.64	2.36E-05	1.30E-03
K00430		00680 Methane metabolism	peroxidase [EC:1.11.1.7]	3	13	-4.19	1.52E-07	1.54E-05
K00123			formate dehydrogenase, alpha subunit [EC:1.2.1.2]	69	2	3.04	6.30E-05	2.97E-03
K00266		00910 Nitrogen metabolism	glutamate synthase (NADPH/NADH) small chain [EC:1.4.1.13] [1.4.1.14]	32	40	-2.39	8.32E-12	2.95E-09
K00265			glutamate synthase (NADPH/NADH) large chain [EC:1.4.1.13] [1.4.1.14]	100	72	-1.60	8.94E-12	2.95E-09
K01914			aspartate--ammonia ligase [EC:6.3.1.1]	0	4	NA	1.37E-03	3.87E-02
K00260			glutamate dehydrogenase [EC:1.4.1.2]	0	5	NA	2.63E-04	8.99E-03
K01424			L-asparaginase [EC:3.5.1.1]	0	4	NA	1.37E-03	3.87E-02
K06445	01130 Lipid Metabolism	00071 Fatty acid metabolism	acyl-CoA dehydrogenase [EC:1.3.99.-]	13	41	-3.73	3.13E-19	2.75E-16
K01897			long-chain acyl-CoA synthetase [EC:6.2.1.3]	21	18	-1.85	1.21E-04	4.92E-03
K01046		00561 Glycerolipid metabolism	triacylglycerol lipase [EC:3.1.1.3]	1	19	-6.32	4.02E-13	2.12E-10
K01755	01150 Amino Acid Metabolism	00220 Urea cycle and metabolism of amino groups	argininosuccinate lyase [EC:4.3.2.1]	12	11	-1.95	1.79E-03	4.77E-02
K00831		00260 Glycine, serine and threonine metabolism	phosphoserine aminotransferase [EC:2.6.1.52]	5	9	-2.92	2.79E-04	9.31E-03
K00003		00300 Lysine biosynthesis	homoserine dehydrogenase [EC:1.1.1.3]	10	13	-2.45	7.93E-05	3.67E-03
K00800		00400 Phenylalanine, tyrosine and tryptophan biosynthesis	3-phosphoshikimate 1-carboxyvinyltransferase [EC:2.5.1.19]	15	13	-1.86	9.87E-04	3.14E-02
K01423	01190 Metabolism of Cofactors and Vitamins	00780 Biotin metabolism	peptidase, M28 (aminopeptidase S) family [EC:3.4.-.-]	12	14	-2.29	8.59E-05	3.84E-03
K03089	01210 Transcription	03020 RNA polymerase	RNA polymerase sigma-32 factor	38	24	-1.41	3.17E-04	1.05E-02
K02965	01220 Translation	03010 Ribosome	small subunit ribosomal protein S19	35	25	-1.59	5.60E-05	2.82E-03
K07566		03014 Other translation proteins	putative translation factor	3	7	-3.29	6.71E-04	2.19E-02
K07576			metallo-beta-lactamase family protein	0	7	NA	9.70E-06	6.40E-04
K02453	01230 Folding, Sorting and Degradation	03090 Type II secretion system	general secretion pathway protein D	2	6	-3.66	9.87E-04	3.14E-02
K04088		03100 Protein folding and associated processing	membrane protease subunit HflK [EC:3.4.-.-]	36	32	-1.90	1.87E-07	1.83E-05
K03111	01240 Replication and Repair	03030 DNA replication	single-strand DNA-binding protein	4	7	-2.88	1.54E-03	4.23E-02
K07493		03034 Other replication, recombination and repair proteins	putative transposase	0	15	NA	1.80E-11	5.29E-09
K07486			transposase	0	9	NA	3.58E-07	2.95E-05
K09969	01310 Membrane Transport	02010 ABC transporters	general L-amino acid transport system substrate-binding protein	141	11	1.61	4.85E-05	2.61E-03
K01999			branched-chain amino acid transport system substrate-binding protein	158	12	1.65	1.38E-05	8.16E-04
K02035			peptide/nickel transport system substrate-binding protein	127	5	2.60	3.35E-07	2.95E-05
K02055			putative spermidine/putrescine transport system substrate-binding protein	111	3	3.14	1.49E-07	1.54E-05
K05559		02052 Other ion-coupled transporters	multicomponent:H antiporter subunit A	1	5	-4.39	1.32E-03	3.87E-02
K03307			solute:Na symporter, SSS family	296	38	0.89	1.59E-04	5.81E-03
K03320			ammonium transporter, Amt family	718	42	2.02	7.40E-28	1.95E-24
K02168			high-affinity choline transport protein	0	12	NA	2.54E-09	4.19E-07
K03286		02070 Pores ion channels	OmpA-OmpF porin, OOP family	5	16	-3.75	2.61E-08	3.12E-06
K04043			molecular chaperone DnaK	130	105	-1.76	7.67E-19	5.06E-16
K02014			iron complex outermembrane recepter protein	184	138	-1.66	2.40E-22	3.16E-19
K07507		02082 Other transporters	putative Mg2 transporter-C (MgtC) family protein	0	4	NA	1.37E-03	3.87E-02
K03413	01320 Signal Transduction	02020 Two-component system	two-component system, chemotaxis family, response regulator CheY	2	11	-4.53	7.01E-07	5.44E-05
K07659			two-component system, OmpR family, phosphate regulon response regulator OmpR	2	8	-4.07	5.77E-05	2.82E-03
K03407			two-component system, chemotaxis family, sensorinase CheA [EC:2.7.13.3]	6	15	-3.39	2.98E-07	2.71E-05
K07806			UDP-4-amino-4-deoxy-L-arabinose-oxoglutarate aminotransferase [EC:2.6.1.-]	4	7	-2.88	1.54E-03	4.23E-02
K07795			putative tricarboxylic transport membrane protein	56	1	3.74	1.36E-04	5.36E-03
K07773			two-component system, OmpR family, aerobic respiration control protein ArcA	0	7	NA	9.70E-06	6.40E-04
K07662			two-component system, OmpR family, response regulator CpxR	0	16	NA	3.47E-12	1.52E-09
K03408	01410 Cell Motility	02030 Bacterial chemotaxis	purine-binding chemotaxis protein CheW	1	8	-5.07	1.39E-05	8.16E-04
K03406			methyl-accepting chemotaxis protein	18	27	-2.66	1.98E-09	3.72E-07
K02391		02040 Flagellar assembly	flagellar basal-body rod protein FlgF	1	8	-5.07	1.39E-05	8.16E-04
K02404			flagellar biosynthesis protein FlhF	5	12	-3.33	5.95E-06	4.13E-04
K02416			flagellar motor switch protein FliM	9	14	-2.71	1.31E-05	8.16E-04
K02407			flagellar hook-associated protein 2	10	12	-2.33	2.37E-04	8.43E-03
K02556			chemotaxis protein MotA	13	14	-2.18	1.47E-04	5.46E-03
K02409			flagellar M-ring protein FliF	10	10	-2.07	1.90E-03	4.96E-02
K02390			flagellar hook protein FlgE	18	17	-1.99	8.44E-05	3.84E-03
K02406			flagellin	260	103	-0.73	2.36E-05	1.30E-03
K02396			flagellar hook-associated protein 1 FlgK	0	12	NA	2.54E-09	4.19E-07
K02414			flagellar hook-length control protein FliK	0	9	NA	3.58E-07	2.95E-05
K02395			flagellar protein FlgJ	0	4	NA	1.37E-03	3.87E-02
K03798	01420 Cell Growth and Death	04410 Cell division	cell division protease FtsH [EC:3.4.24.-]	209	85	-0.77	5.70E-05	2.82E-03

*KO = KEGG ortholog number; Control and DOM are raw counts of sequences annotated as a KEGG ORF in the controls and treatments. Note for this analysis, all controls were pooled based on the results of the ANOVA in Supplemental Table; ln(Fold change) is the natural log of the estimated fold change of the pooled controls relative to the treatment (i.e., positive values indicate enrichment in the controls). The fold changes are calculated after scaling by the number of non-rRNA reads in the library (see Methods); *q*-value is a calibration of the table-wide false discovery rate (Storey et al., 2003). Note that some KEGG othologs belong to multiple functional categories and pathways. For brevity, we have included only one designation for each ortholog.

Table S5 : Pairwise tests of Functional Annotations Between Controls (Pooled) and HMWDOM Treatments 12 Hours Post Addition *

KO	Functional category	Pathway	ORF Annotation	Control	DOM12	ln(Fold change)	p-value	q-value
K00134	01110 Carbohydrate Metabolism	00010 Glycolysis / Gluconeogenesis	glyceraldehyde 3-phosphate dehydrogenase [EC:1.2.1.12]	30	28	-1.88	1.66E-06	4.71E-05
K01803			triosephosphate isomerase (TIM) [EC:5.3.1.1]	8	9	-2.15	2.82E-03	2.73E-02
K00164		00020 Citrate cycle (TCA cycle)	2-oxoglutarate dehydrogenase E1 component [EC:1.2.4.2]	22	18	-1.69	4.73E-04	5.44E-03
K00658			2-oxoglutarate dehydrogenase E2 component (dihydrolipoamide succinyltransferase) [EC:2.3.1.61]	9	14	-2.61	2.43E-05	4.70E-04
K01681			aconitate hydratase 1 [EC:4.2.1.3]	29	23	-1.64	8.66E-05	1.29E-03
K01682			aconitate hydratase 2 [EC:4.2.1.3]	51	50	-1.95	5.27E-11	2.87E-09
K01647			citrate synthase [EC:2.3.3.1]	19	17	-1.82	2.54E-04	3.43E-03
K01676			fumarate hydratase, class I [EC:4.2.1.2]	4	10	-3.30	5.15E-05	8.53E-04
K01902			succinyl-CoA synthetase alpha subunit [EC:6.2.1.5]	19	22	-2.19	2.29E-06	6.16E-05
K01903			succinyl-CoA synthetase beta subunit [EC:6.2.1.5]	28	18	-1.34	2.93E-03	2.78E-02
K00033		00030 Pentose phosphate pathway	6-phosphogluconate dehydrogenase [EC:1.1.1.44]	4	10	-3.30	5.15E-05	8.53E-04
K00948			ribose-phosphate pyrophosphokinase [EC:2.7.6.1]	24	18	-1.56	8.07E-04	9.08E-03
K08093		00040 Pentose and glucuronate interconversions	3-hexulose-6-phosphate synthase [EC:4.1.2.-]	5	10	-2.98	1.26E-04	1.85E-03
K01199		00500 Starch and sucrose metabolism	glucan endo-1,3-beta-D-glucosidase [EC:3.2.1.39]	0	8	NA	2.82E-06	7.38E-05
K01443		00530 Aminosugars metabolism	N-acetylglucosamine-6-phosphate deacetylase [EC:3.5.1.25]	0	4	NA	1.68E-03	1.70E-02
K03431			phosphoglucosamine mutase [EC:5.4.2.10]	13	11	-1.74	4.19E-03	3.78E-02
K00790			UDP-N-acetylglucosamine 1-carboxyvinyltransferase [EC:2.5.1.7]	3	17	-4.48	9.72E-10	4.32E-08
K03777		00620 Pyruvate metabolism	D-lactate dehydrogenase [EC:1.1.1.28]	1	5	-4.30	1.70E-03	1.70E-02
K00027			malate dehydrogenase (oxaloacetate-decarboxylating) [EC:1.1.1.38]	0	4	NA	1.68E-03	1.70E-02
K00026			malate dehydrogenase [EC:1.1.1.37]	20	33	-2.70	3.27E-11	1.89E-09
K01638			malate synthase [EC:2.3.3.9]	27	38	-2.47	1.68E-11	1.09E-09
K01571			oxaloacetate decarboxylase, alpha subunit [EC:4.1.1.3]	5	14	-3.46	7.95E-07	2.35E-05
K01572			oxaloacetate decarboxylase, beta subunit [EC:4.1.1.3]	6	13	-3.09	7.50E-06	1.72E-04
K01610			phosphoenolpyruvate carboxykinase (ATP) [EC:4.1.1.49]	20	46	-3.18	5.91E-18	1.21E-15
K01007			pyruvate,water dikinase [EC:2.7.9.2]	2	104	-7.68	2.02E-69	2.69E-66
K01637		00630 Glyoxylate and dicarboxylate metabolism	isocitrate lyase [EC:4.1.3.1]	62	94	-2.58	1.16E-27	6.17E-25
K01966		00640 Propanoate metabolism	propionyl-CoA carboxylase beta chain [EC:6.4.1.3]	25	0	NA	5.21E-03	4.37E-02
K01652		00650 Butanoate metabolism	acetolactate synthase I/II/III large subunit [EC:2.2.1.6]	60	37	-1.28	6.25E-05	9.85E-04
K01653			acetolactate synthase I/III small subunit [EC:2.2.1.6]	9	9	-1.98	4.64E-03	4.03E-02
K00404	01120 Energy Metabolism	00190 Oxidative phosphorylation	cb-type cytochrome c oxidase subunit I [EC:1.9.3.1]	0	8	NA	2.82E-06	7.38E-05
K00405			cb-type cytochrome c oxidase subunit II [EC:1.9.3.1]	2	17	-5.06	1.80E-10	8.73E-09
K00406			cb-type cytochrome c oxidase subunit III [EC:1.9.3.1]	8	13	-2.68	3.72E-05	6.62E-04
K00425			cytochrome bd-I oxidase subunit I [EC:1.10.3.-]	0	4	NA	1.68E-03	1.70E-02
K00426			cytochrome bd-I oxidase subunit II [EC:1.10.3.-]	2	7	-3.78	3.40E-04	4.24E-03
K02274			cytochrome c oxidase subunit I [EC:1.9.3.1]	197	22	1.19	6.75E-05	1.05E-03
K01507			inorganic pyrophosphatase [EC:3.6.1.1]	170	12	1.85	4.13E-07	1.36E-05
K03885			NADH dehydrogenase [EC:1.6.99.3]	5	7	-2.46	4.20E-03	3.78E-02
K01533		00192 ATPases	Cu2-exporting ATPase [EC:3.6.3.4]	9	10	-2.13	1.74E-03	1.73E-02
K02111		00195 Photosynthesis	F-type H-transporting ATPase subunit alpha [EC:3.6.3.14]	187	85	-0.84	1.99E-05	3.95E-04
K02112			F-type H-transporting ATPase subunit beta [EC:3.6.3.14]	140	102	-1.52	8.92E-15	1.32E-12
K02113			F-type H-transporting ATPase subunit delta [EC:3.6.3.14]	51	28	-1.11	1.76E-03	1.75E-02
K02690			photosystem I core protein Ib	627	75	1.09	1.49E-11	1.04E-09
K02694			photosystem I subunit III	35	0	NA	5.28E-04	6.04E-03
K02703			photosystem II PsbA protein	560	66	1.11	8.71E-11	4.55E-09
K02704			photosystem II PsbB protein	235	30	0.99	1.61E-04	2.26E-03
K02705			photosystem II PsbC protein	352	47	0.93	1.10E-05	2.46E-04
K02706			photosystem II PsbD protein	195	20	1.31	1.85E-05	3.74E-04
K01601		00710 Carbon fixation	ribulose-bisphosphate carboxylase large chain [EC:4.1.1.39]	439	14	2.99	2.61E-27	9.94E-25
K01602			ribulose-bisphosphate carboxylase small chain [EC:4.1.1.39]	90	8	1.51	1.55E-03	1.63E-02
K02198		00910 Nitrogen metabolism	cytochrome c-type biogenesis protein CcmF	10	13	-2.36	1.39E-04	2.02E-03
K00260			glutamate dehydrogenase [EC:1.4.1.2]	0	5	NA	3.40E-04	4.24E-03
K00265			glutamate synthase (NADPH/NADH) large chain [EC:1.4.1.13] [1.4.1.14]	100	97	-1.93	1.45E-19	4.30E-17
K00266			glutamate synthase (NADPH/NADH) small chain [EC:1.4.1.13] [1.4.1.14]	32	36	-2.15	2.45E-09	1.04E-07
K01915			glutamine synthetase [EC:6.3.1.2]	76	41	-1.09	1.91E-04	2.63E-03
K00362			nitrite reductase (NAD(P)H) large subunit [EC:1.7.1.4]	0	4	NA	1.68E-03	1.70E-02
K00390		00920 Sulfur metabolism	phosphoadenosine phosphosulfate reductase [EC:1.8.4.8]	3	10	-3.71	1.80E-05	3.66E-04
K00640			serine O-acetyltransferase [EC:2.3.1.30]	3	6	-2.98	3.28E-03	3.05E-02
K00380			sulfite reductase (NADPH) flavoprotein alpha-component [EC:1.8.1.2]	2	15	-4.88	3.51E-09	1.46E-07
K00381			sulfite reductase (NADPH) hemoprotein beta-component [EC:1.8.1.2]	16	31	-2.93	1.43E-11	1.03E-09
K00647	01130 Lipid Metabolism	00061 Fatty acid biosynthesis	3-oxoacyl-[acyl-carrier-protein] synthase I [EC:2.3.1.41]	6	13	-3.09	7.50E-06	1.72E-04
K00022			3-hydroxyacyl-CoA dehydrogenase [EC:1.1.1.35]	47	56	-2.23	2.26E-14	2.31E-12
K00632			acetyl-CoA acyltransferase [EC:2.3.1.16]	26	28	-2.08	2.53E-07	8.44E-06
K06445			acyl-CoA dehydrogenase [EC:1.3.99.-]	13	100	-4.92	7.32E-55	6.50E-52
K00631		00561 Glycerolipid metabolism	glycerol-3-phosphate O-acyltransferase [EC:2.3.1.15]	1	6	-4.56	3.99E-04	4.77E-03
K00951	01140 Nucleotide Metabolism	00230 Purine metabolism	GTP pyrophosphokinase [EC:2.7.6.5]	10	12	-2.24	3.94E-04	4.77E-03
K00940			nucleoside-diphosphatekinase [EC:2.7.4.6]	12	15	-2.30	5.58E-04	6.09E-04
K01945			phosphoribosylamine--glycine ligase [EC:6.3.4.13]	16	14	-1.78	1.02E-03	1.13E-02
K01589			phosphoribosylaminoimidazole carboxylase ATPase subunit [EC:4.1.1.21]	6	12	-2.98	2.56E-05	4.80E-04
K01933			phosphoribosylformylglycinamidine cyclo-ligase [EC:6.3.3.1]	5	9	-2.83	4.21E-04	4.99E-03
K01952			phosphoribosylformylglycinamidine synthase [EC:6.3.5.3]	21	22	-2.04	6.44E-06	1.53E-04
K00945		00240 Pyrimidine metabolism	cytidylatekinase [EC:2.7.4.14]	8	12	-2.56	1.15E-04	1.70E-03
K01755	01150 Amino Acid Metabolism	00220 Urea cycle and metabolism of amino groups	argininosuccinate lyase [EC:4.3.2.1]	12	14	-2.20	1.55E-04	2.21E-03
K01955		00251 Glutamate metabolism	carbamoyl-phosphate synthase large subunit [EC:6.3.5.5]	36	28	-1.61	2.65E-05	4.90E-04
K01580			glutamate decarboxylase [EC:4.1.1.15]	0	5	NA	3.40E-04	4.24E-03
K01939		00252 Alanine and aspartate metabolism	adenylosuccinate synthase [EC:6.3.4.4]	10	16	-2.66	5.16E-06	1.26E-04
K01940			argininosuccinate synthase [EC:6.3.4.5]	13	16	-2.28	3.52E-05	6.39E-04
K01876			aspartyl-tRNA synthetase [EC:6.1.1.12]	20	14	-1.46	4.79E-03	4.13E-02
K00315		00260 Glycine, serine and threonine metabolism	dimethylglycine dehydrogenase [EC:1.5.99.2]	38	1	3.27	2.41E-03	2.38E-02
K01879			glycyl-tRNA synthetase beta chain [EC:6.1.1.14]	10	14	-2.46	4.75E-05	8.17E-04
K00831			phosphoserine aminotransferase [EC:2.6.1.52]	5	11	-3.11	3.68E-05	6.58E-04
K01754			threonine dehydratase [EC:4.3.1.19]	6	12	-2.98	2.56E-05	4.80E-04
K01740		00271 Methionine metabolism	O-acetylhomoserine (thiol)-lyase [EC:2.5.1.49]	119	9	1.75	5.83E-05	9.42E-04
K00166		00280 Valine, leucine and isoleucine degradation	2-oxoisovalerate dehydrogenase E1 component, alpha subunit [EC:1.2.4.4]	0	4	NA	1.68E-03	1.70E-02
K00053			ketol-acid reductoisomerase [EC:1.1.1.86]	36	30	-1.71	3.79E-06	9.54E-05
K00821		00300 Lysine biosynthesis	acetylornithine/N-succinyldiaminopimelate aminotransferase [EC:2.6.1.11] [2.6.1.17]	15	20	-2.39	1.77E-06	4.97E-05
K00003			homoserine dehydrogenase [EC:1.1.1.3]	10	20	-2.98	4.74E-06	1.86E-05
K00673		00330 Arginine and proline metabolism	arginine N-succinyltransferase [EC:2.3.1.109]	2	8	-3.98	8.55E-05	1.29E-03
K01887			arginyl-tRNA synthetase [EC:6.1.1.19]	7	8	-2.17	4.59E-03	4.03E-02
K06447			succinylglutamic semialdehyde dehydrogenase [EC:1.2.1.71]	1	7	-4.78	9.19E-05	1.36E-03
K01892		00340 Histidine metabolism	histidyl-tRNA synthetase [EC:6.1.1.21]	9	15	-2.71	7.80E-06	1.76E-04
K00800		00400 Phenylalanine, tyrosine and tryptophan biosynthesis	3-phosphoshikimate 1-carboxyvinyltransferase [EC:2.5.1.19]	15	17	-2.16	3.86E-05	6.73E-04
K01918	01160 Metabolism of Other Amino Acids	00410 beta-Alanine metabolism	pantoate--beta-alanine ligase [EC:6.3.2.1]	2	6	-3.56	1.32E-03	1.41E-02
K00979	01170 Glycan Biosynthesis and Metabolism	00540 Lipopolysaccharide biosynthesis	3-deoxy-manno-octulosonate cytidylyltransferase (CMP-KDO synthetase) [EC:2.7.7.38]	5	8	-2.66	1.36E-03	1.44E-02
K02535			UDP-3-O-[3-hydroxymyristoyl] N-acetylglucosamine deacetylase [EC:3.5.1.-]	20	19	-1.90	6.55E-05	1.03E-03
K01925		00550 Peptidoglycan biosynthesis	UDP-N-acetylmuramoylalanine--D-glutamate ligase [EC:6.3.2.9]	9	9	-1.98	4.64E-03	4.03E-02
K01929			UDP-N-acetylmuramoylalanyl-D-glutamyl-2, 6-diaminopimelate--D-alanyl-D-alanine ligase [EC:6.3.2.10]	5	7	-2.46	4.20E-03	3.78E-02
K06153			undecaprenyl-diphosphatase [EC:3.6.1.27]	2	5	-3.30	4.95E-03	4.17E-02
K00602	01190 Metabolism of Cofactors and Vitamins	00670 One carbon pool by folate	phosphoribosylaminoimidazolecarboxamide formyltransferase [EC:2.1.2.3]	7	14	-2.98	5.24E-06	1.26E-04
K00560			thymidylate synthase [EC:2.1.1.45]	3	8	-3.39	2.58E-04	3.43E-03
K00946		00730 Thiamine metabolism	thiamine monophosphatekinase [EC:2.7.4.16]	3	6	-2.98	3.28E-03	3.05E-02
K00794		00740 Riboflavin metabolism	riboflavin synthase beta chain [EC:2.5.1.-]	2	8	-3.98	8.55E-05	1.29E-03
K00324		00760 Nicotinate and nicotinamide metabolism	NAD(P) transhydrogenase subunit alpha [EC:1.6.1.2]	27	40	-2.54	1.93E-12	1.71E-10
K00325			NAD(P) transhydrogenase subunit beta [EC:1.6.1.2]	29	52	-2.82	1.09E-17	1.94E-15
K03517			quinolinate synthase	7	9	-2.34	1.62E-03	1.69E-02
K01737		00790 Folate biosynthesis	6-pyruvoyl tetrahydrobiopterin synthase [EC:4.2.3.12]	42	1	3.41	1.04E-03	1.14E-02
K00228		00860 Porphyrin and chlorophyll metabolism	coproporphyrinogen III oxidase [EC:1.3.3.3]	2	5	-3.30	4.95E-03	4.17E-02
K02302			uroporphyrin-III C-methyltransferase [EC:2.1.1.107]	5	7	-2.46	4.20E-03	3.78E-02
K01835	01195 Biosynthesis of Secondary Metabolites	00521 Streptomycin biosynthesis	phosphoglucomutase [EC:5.4.2.2]	1	8	-4.98	2.08E-05	4.09E-04
K03040	01210 Transcription	03020 RNA polymerase	DNA-directed RNA polymerase subunit alpha [EC:2.7.7.6]	91	65	-1.49	1.32E-09	5.69E-08
K02600		03028 Other transcription related proteins	N utilization substance protein A	48	26	-1.09	3.29E-03	3.06E-02
K03628			transcription termination factor Rho	14	14	-1.98	4.26E-04	5.02E-03
K02502	01220 Translation	00970 Aminoacyl-tRNA biosynthesis	ATP phosphonbosyltransferase regulatory subunit	2	5	-3.30	4.95E-03	4.17E-02
K02863		03010 Ribosome	large subunit ribosomal protein L1	94	46	-0.95	4.56E-04	5.31E-03
K02867			large subunit ribosomal protein L11	30	20	-1.39	1.26E-03	1.37E-02
K02871			large subunit ribosomal protein L13	33	20	-1.25	3.15E-03	2.98E-02
K02874			large subunit ribosomal protein L14	30	20	-1.39	1.26E-03	1.37E-02
K02876			large subunit ribosomal protein L15	38	30	-1.64	1.11E-05	2.46E-04
K02878			large subunit ribosomal protein L16	49	40	-1.68	1.71E-07	6.32E-06
K02881			large subunit ribosomal protein L18	44	34	-1.61	3.02E-06	7.83E-05
K02886			large subunit ribosomal protein L2	95	47	-0.96	3.32E-04	4.24E-03
K02890			large subunit ribosomal protein L22	29	24	-1.70	4.09E-05	7.12E-04
K02892			large subunit ribosomal protein L23	28	19	-1.42	1.55E-03	1.63E-02
K02897			large subunit ribosomal protein L25	24	29	-2.25	3.25E-08	1.29E-06
K02906			large subunit ribosomal protein L3	71	62	-1.78	1.07E-11	8.11E-10

Table S5 : Pairwise tests of Functional Annotations Between Controls (Pooled) and HMWDOM Treatments 12 Hours Post Addition * (cont)

KO	Functional category	Pathway	ORF Annotation	Control	DOM12	ln(Fold change)	p-value	q-value
K02907			large subunit ribosomal protein L30	6	13	-3.09	7.50E-06	1.72E-04
K02911			large subunit ribosomal protein L32	21	35	-2.71	7.09E-12	5.90E-10
K02919			large subunit ribosomal protein L36	8	15	-2.88	3.60E-06	9.24E-05
K02926			large subunit ribosomal protein L4	64	61	-1.91	1.46E-12	1.34E-10
K02931			large subunit ribosomal protein L5	75	45	-1.24	1.17E-05	2.58E-04
K02933			large subunit ribosomal protein L6	44	46	-2.04	7.63E-11	4.07E-09
K02945			small subunit ribosomal protein S1	98	47	-0.92	5.74E-04	6.51E-03
K02948			small subunit ribosomal protein S11	69	47	-1.42	6.25E-07	1.89E-05
K02950			small subunit ribosomal protein S12	45	29	-1.34	2.10E-04	2.87E-03
K02954			small subunit ribosomal protein S14	37	41	-2.13	2.61E-10	1.22E-08
K02963			small subunit ribosomal protein S18	17	13	-1.59	4.61E-03	4.03E-02
K02965			small subunit ribosomal protein S19	35	21	-1.24	2.53E-03	2.49E-02
K02982			small subunit ribosomal protein S3	99	74	-1.56	1.65E-11	1.09E-09
K02986			small subunit ribosomal protein S4	87	55	-1.32	5.06E-07	1.61E-05
K02988			small subunit ribosomal protein S5	66	55	-1.71	4.14E-10	1.87E-08
K02990			small subunit ribosomal protein S6	16	14	-1.78	1.02E-03	1.13E-02
K02860		03014 Other translation proteins	16S rRNA processing protein RimM	20	34	-2.74	1.04E-11	8.11E-10
K02528			dimethyladenosine transferase [EC:2.1.1.-]	1	6	-4.56	3.99E-04	4.77E-03
K01265			methionyl aminopeptidase [EC:3.4.11.18]	4	9	-3.15	1.83E-04	2.57E-03
K07320			putative adenine-specific DNA-methyltransferase [EC:2.1.1.72]	0	5	NA	3.40E-04	4.24E-03
K03685			ribonuclease III [EC:3.1.26.3]	5	7	-2.46	4.20E-03	3.78E-02
K06178			ribosomal large subunit pseudouridine synthase B [EC:5.4.99.12]	4	10	-3.30	5.15E-05	8.53E-04
K02687			ribosomal protein L11 methyltransferase [EC:2.1.1.-]	1	6	-4.56	3.99E-04	4.77E-03
K03218			RNA methyltransferase, TrmH family [EC:2.1.1.-]	2	6	-3.56	1.32E-03	1.41E-02
K00554			tRNA (guanine-N1-)-methyltransferase [EC:2.1.1.31]	28	48	-2.76	5.07E-16	8.44E-14
K02533			tRNA/rRNA methyltransferase [EC:2.1.1.-]	1	5	-4.30	1.70E-03	1.70E-02
K05540			tRNA-dihydrouridine synthase B [EC:1.-.-.-]	12	14	-2.20	1.55E-04	2.21E-03
K03076	01230 Folding, Sorting and Degradation	03060 Protein export	preprotein translocase SecY subunit	75	115	-2.59	9.40E-34	6.26E-31
K03106			signal recognition particle, subunit SRP54	19	14	-1.54	3.69E-03	3.42E-02
K02453		03090 Type II secretion system	general secretion pathway protein D	2	5	-3.30	4.95E-03	4.17E-02
K02455			general secretion pathway protein F	0	6	NA	6.89E-05	1.06E-03
K02666			type IV pilus assembly protein PilQ	2	5	-3.30	4.95E-03	4.17E-02
K01802		03100 Protein folding and associated processing	peptidylprolyl isomerase [EC:5.2.1.8]	34	1	3.11	5.58E-03	4.66E-02
K05838			putative thioredoxin	2	7	-3.78	3.40E-04	4.24E-03
K03580	01240 Replication and Repair	03034 Other replication, recombination and repair proteins	ATP-dependent helicase HepA [EC:3.6.1.-]	0	8	NA	2.82E-06	7.38E-05
K05592			ATP-dependent RNA helicase DeaD	19	16	-1.73	9.68E-04	1.08E-02
K03732			ATP-dependent RNA helicase RhlB [EC:3.6.1.-]	2	6	-3.56	1.32E-03	1.41E-02
K04485			DNA repair protein RadA/Sms	0	4	NA	1.68E-03	1.70E-02
K03582			exodeoxyribonuclease V beta subunit [EC:3.1.11.5]	0	6	NA	6.89E-05	1.06E-03
K07493			putative transposase	0	7	NA	1.40E-05	2.86E-04
K07497			putative transposase	1	6	-4.56	3.99E-04	4.77E-03
K00986			RNA-directed DNA polymerase [EC:2.7.7.49]	3	6	-2.98	3.28E-03	3.05E-02
K04047			starvation-inducible DNA-binding protein	2	5	-3.30	4.95E-03	4.17E-02
K07486			transposase	0	10	NA	1.16E-07	4.34E-06
K01999	01310 Membrane Transport	02010 ABC transporters	branched-chain amino acid transport system substrate-binding protein	158	6	2.74	1.07E-09	4.67E-08
K09969			general L-amino acid transport system substrate-binding protein	141	2	4.16	8.90E-12	7.19E-10
K02011			iron(III) transport system permease protein	2	7	-3.78	3.40E-04	4.24E-03
K02035			peptide/nickel transport system substrate-binding protein	127	5	2.69	8.92E-08	3.45E-06
K02037			phosphate transport system permease protein	8	11	-2.44	3.47E-04	4.30E-03
K02038			phosphate transport system permease protein	3	24	-4.98	3.42E-14	3.37E-12
K02040			phosphate transport system substrate-binding protein	32	25	-1.62	4.90E-05	8.37E-04
K02065			putative ABC transport system ATP-binding protein	5	7	-2.46	4.20E-03	3.78E-02
K02055			putative spermidine/putrescine transport system substrate-binding protein	111	1	4.82	4.07E-10	1.87E-08
K02051			sulfonate/nitrate/taurine transport system substrate-binding protein	52	2	2.72	9.92E-04	1.11E-02
K03320		02052 Other ion-coupled transporters	ammonium transporter, Amt family	718	49	1.90	3.39E-27	1.13E-24
K03317			concentrative nucleoside transporter, CNT family	2	5	-3.30	4.95E-03	4.17E-02
K07301			ecm27	5	7	-2.46	4.20E-03	3.78E-02
K07239			heavy-metal exporter, HME family	3	8	-3.39	2.58E-04	3.43E-03
K02168			high-affinity choline transport protein	0	10	NA	1.16E-07	4.34E-06
K03296			hydrophobic/amphiphilic exporter-1 (mainly G- bacteria), HAE1 family	16	14	-1.78	1.02E-03	1.13E-02
K03306			inorganic phosphate transporter, PiT family	6	12	-2.98	2.56E-05	4.80E-04
K03315			Na:H antiporter, NhaC family	0	9	NA	5.72E-07	1.75E-05
K03308			neurotransmitter:Na symporter, NSS family	7	10	-2.49	5.47E-04	6.23E-03
K03307			solute:Na symporter, SSS family	296	23	1.71	1.77E-10	8.73E-09
K08484		02060 Phosphotransferase system (PTS)	phosphotransferase system, enzyme I, PtsP [EC:2.7.3.9]	1	11	-5.44	2.29E-07	8.36E-06
K02014		02070 Pores ion channels	iron complex outermembrane recepter protein	184	134	-1.52	7.23E-19	1.93E-16
K03286			OmpA-OmpF porin, OOP family	5	23	-4.18	3.70E-12	3.18E-10
K03287			outer membrane factor, OMF family	5	8	-2.66	1.36E-03	1.44E-02
K08720			outer membrane protein OmpU	51	0	NA	1.39E-05	2.86E-04
K01991			polysaccharide export outer membrane protein	9	10	-2.13	1.74E-03	1.73E-02
K03811		02082 Other transporters	nicotinamide mononucleotide transporter	0	7	NA	1.40E-05	2.86E-04
K03412	01320 Signal Transduction	02020 Two-component system	protein-glutamate methylesterase, two-component system, chemotaxis family, response regulator CheB [EC:3.1.1.61]	0	9	NA	5.72E-07	1.75E-05
K07795			putative tricarboxylic transport membrane protein	56	2	2.83	4.41E-04	5.15E-03
K03415			two-component system, chemotaxis family, response regulator CheV	0	7	NA	1.40E-05	2.86E-04
K03413			two-component system, chemotaxis family, response regulator CheY	2	17	-5.06	1.80E-10	8.73E-09
K03407			two-component system, chemotaxis family, sensorinase CheA [EC:2.7.13.3]	6	18	-3.56	1.23E-08	4.95E-07
K07675			two-component system, NarL family, sensor histidinekinase UhpB [EC:2.7.13.3]	0	7	NA	1.40E-05	2.86E-04
K07715			two-component system, NtrC family, response regulator YfhA	7	14	-2.98	5.24E-06	1.26E-04
K07711			two-component system, NtrC family, sensor histidinekinase YfhK [EC:2.7.13.3]	0	7	NA	1.40E-05	2.86E-04
K07773			two-component system, OmpR family, aerobic respiration control protein ArcA	0	5	NA	3.40E-04	4.24E-03
K07657			two-component system, OmpR family, phosphate regulon response regulator PhoB	2	26	-5.78	1.05E-17	1.94E-15
K07636			two-component system, OmpR family, phosphate regulon sensor histidinekinase PhoR [EC:2.7.13.3]	0	19	NA	6.59E-14	6.27E-12
K07662			two-component system, OmpR family, response regulator CpxR	0	7	NA	1.40E-05	2.86E-04
K02488			two-component system, PleD related family, response regulator	4	10	-3.30	5.15E-05	8.53E-04
K00575	01410 Cell Motility	02030 Bacterial chemotaxis	chemotaxis protein methyltransferase CheR [EC:2.1.1.80]	2	8	-3.98	8.55E-05	1.29E-03
K03406			methyl-accepting chemotaxis protein	18	30	-2.71	2.24E-10	1.06E-08
K03408			purine-binding chemotaxis protein CheW	1	8	-4.98	1.08E-05	4.09E-04
K02404		02040 Flagellar assembly	flagellar biosynthesis protein FlhF	5	9	-2.83	4.21E-04	4.99E-03
K02416			flagellar motor switch protein FliM	9	9	-1.98	4.64E-03	4.03E-02
K02409			flagellar M-ring protein FliF	10	12	-2.24	3.94E-04	4.77E-03
K02395			flagellar protein FlgJ	0	4	NA	1.68E-03	1.70E-02
K03590	01420 Cell Growth and Death	04410 Cell division	cell division protein FtsA	6	8	-2.39	1.60E-03	2.55E-02
K03496			chromosome partitioning protein	10	10	-1.98	2.86E-03	2.75E-02
K03529			chromosome segregation protein	15	17	-2.16	3.86E-05	6.76E-04
K03495			glucose inhibited division protein A	23	20	-1.78	1.56E-04	2.21E-03
K03570			rod shape-determining protein MreC	1	5	-4.30	1.70E-03	1.70E-02
K03609			septum site-determining protein MinD	0	4	NA	1.68E-03	1.70E-02

KO = KEGG ortholog number; Control and DOM are raw counts of sequences annotated as a KEGG ORF in the controls and treatments. Note for this analysis, all controls were pooled based on the results of the ANOVA in Supplemental Table; ln(Fold change) is the natural log of the estimated fold change of the pooled controls relative to the treatment (i.e., positive values indicate enrichment in the controls). The fold changes are calculated after scaling by the number of non-rRNA reads in the library (see Methods); *q*-value is a calibration of the table-wide false discovery rate (Storey et al., 2003). Note that some KEGG othologs belong to multiple functional categories and pathways. For brevity, we have included only one designation for each ortholog.

Supplemental Table S6 : Pairwise tests of Functional Annotations Between Controls (Pooled) and HMWDOM Treatments 27 Hours Post Addition *

KO	Functional category	Pathway	ORF Annotation	Control	DOM27	ln(Fold change)	p-value	q-value
K01803	01110 Carbohydrate Metabolism	00010 Glycolysis / Gluconeogenesis	triosephosphate isomerase (T/M) [EC:5.3.1.1]	8	45	-3.03	7.67E-11	1.73E-09
K01810			glucose-6-phosphate isomerase [EC:5.3.1.9]	5	17	-2.30	7.16E-04	5.32E-03
K00134			glyceraldehyde 3-phosphate dehydrogenase [EC:1.2.1.12]	30	63	-1.61	2.08E-07	3.07E-06
K00163			pyruvate dehydrogenase E1 component [EC:1.2.4.1]	20	36	-1.39	5.43E-04	4.15E-03
K00382			dihydrolipoamide dehydrogenase [EC:1.8.1.4]	34	43	-0.88	1.02E-02	4.89E-02
K00161			pyruvate dehydrogenase E1 component subunit alpha [EC:1.2.4.1]	27	3	2.63	3.03E-04	2.47E-03
K02446			fructose-1,6-bisphosphatase II [EC:3.1.3.11]	21	0	NA	1.98E-05	2.02E-04
K00026		00020 Citrate cycle (TCA cycle)	malate dehydrogenase [EC:1.1.1.37]	20	101	-2.88	5.11E-22	4.71E-20
K00658			2-oxoglutarate dehydrogenase E2 component (dihydrolipoamide succinyltransferase) [EC:2.3.1.61]	9	24	-1.95	2.65E-04	2.20E-03
K01647			citrate synthase [EC:2.3.3.1]	19	46	-1.81	1.47E-06	1.80E-05
K01902			succinyl-CoA synthetase alpha subunit [EC:6.2.1.5]	19	44	-1.75	4.04E-06	4.70E-05
K01682			aconitate hydratase 2 [EC:4.2.1.3]	51	78	-1.15	8.67E-06	9.30E-05
K01903			succinyl-CoA synthetase beta subunit [EC:6.2.1.5]	28	39	-1.02	4.15E-03	2.36E-02
K01644			citrate lyase subunit beta [EC:4.1.3.6]	14	1	3.27	6.73E-03	3.52E-02
K01690		00030 Pentose phosphate pathway	phosphogluconate dehydratase [EC:4.2.1.12]	6	38	-3.20	6.24E-10	1.32E-08
K00948			ribose-phosphate pyrophosphokinase [EC:2.7.6.1]	24	46	-1.48	4.19E-05	4.15E-04
K00616			transaldolase [EC:2.2.1.2]	21	39	-1.43	1.85E-04	1.59E-03
K00117			quinoprotein glucose dehydrogenase [EC:1.1.5.2]	26	5	1.84	5.30E-03	2.83E-02
K08093		00040 Pentose and glucuronate interconversions	3-hexulose-6-phosphate synthase [EC:4.1.2..]	5	79	-4.52	3.73E-25	5.74E-23
K08094			6-phospho-3-hexuloisomerase [EC:5.-.-.-]	0	93	NA	5.46E-37	1.37E-34
K01179		00500 Starch and sucrose metabolism	endoglucanase [EC:3.2.1.4]	0	13	NA	8.59E-06	9.30E-05
K00963		00520 Nucleotide sugars metabolism	UTP--glucose-1-phosphate uridylyltransferase [EC:2.7.7.9]	5	18	-2.39	3.85E-04	3.05E-03
K01654		00530 Aminosugars metabolism	N-acetylneuraminate synthase [EC:2.5.1.56]	2	15	-3.45	7.42E-05	6.92E-04
K00983			N-acylneuraminate cytidylyltransferase [EC:2.7.7.43]	4	18	-2.71	1.11E-04	1.00E-03
K00790			UDP-N-acetylglucosamine 1-carboxyvinyltransferase [EC:2.5.1.7]	3	12	-2.54	2.73E-03	1.65E-02
K01007		00620 Pyruvate metabolism	pyruvate,water dikinase [EC:2.7.9.2]	2	173	-6.97	1.71E-64	9.49E-62
K03777			D-lactate dehydrogenase [EC:1.1.1.28]	1	20	-4.86	2.06E-07	3.07E-06
K01571			oxaloacetate decarboxylase, alpha subunit [EC:4.1.1.3]	5	24	-2.80	4.68E-06	5.34E-05
K01638			malate synthase [EC:2.3.3.9]	27	103	-2.47	6.01E-19	4.16E-17
K01610			phosphoenolpyruvate carboxykinase (ATP) [EC:4.1.1.49]	20	73	-2.41	1.91E-13	5.81E-12
K01595			phosphoenolpyruvate carboxylase [EC:4.1.1.31]	5	16	-2.22	1.33E-03	8.97E-03
K01572			oxaloacetate decarboxylase, beta subunit [EC:4.1.1.3]	6	16	-1.95	3.68E-03	2.18E-02
K01649			2-isopropylmalate synthase [EC:2.3.3.13]	37	47	-0.88	5.35E-03	2.84E-02
K01006			pyruvate,orthophosphate dikinase [EC:2.7.9.1]	24	1	4.05	4.90E-05	4.80E-04
K01637		00630 Glyoxylate and dicarboxylate metabolism	isocitrate lyase [EC:4.1.3.1]	62	227	-2.41	3.14E-39	8.69E-37
K00822		00640 Propanoate metabolism	beta-alanine--pyruvate transaminase [EC:2.6.1.18]	19	3	2.12	8.65E-03	4.33E-02
K00925			acetateinase [EC:2.7.2.1]	0	10	NA	1.27E-04	1.14E-03
K01653		00650 Butanoate metabolism	acetolactate synthase I/III small subunit [EC:2.2.1.6]	9	21	-1.76	1.38E-03	9.22E-03
K01652			acetolactate synthase I/II/III large subunit [EC:2.2.1.6]	60	102	-1.30	1.35E-08	2.47E-07
K00135			succinate-semialdehyde dehydrogenase (NADP) [EC:1.2.1.16]	27	2	3.22	8.07E-05	7.50E-04
K00405	01120 Energy Metabolism	00190 Oxidative phosphorylation	cb-type cytochrome c oxidase subunit II [EC:1.9.3.1]	2	28	-4.35	1.96E-09	3.93E-08
K00937			polyphosphateinase [EC:2.7.4.1]	5	29	-3.07	1.32E-07	2.05E-06
K00406			cb-type cytochrome c oxidase subunit III [EC:1.9.3.1]	8	23	-2.06	3.18E-04	2.58E-03
K02275			cytochrome c oxidase subunit II [EC:1.9.3.1]	44	9	1.75	3.65E-04	2.91E-03
K02274			cytochrome c oxidase subunit I [EC:1.9.3.1]	197	35	1.95	2.95E-17	1.67E-15
K01507			inorganic pyrophosphatase [EC:3.6.1.1]	170	21	2.48	1.57E-19	1.17E-17
K00336			NADH dehydrogenase I subunit G [EC:1.6.5.3]	21	2	2.85	1.09E-03	7.53E-03
K00333			NADH dehydrogenase I subunit D [EC:1.6.5.3]	15	1	3.37	3.92E-03	2.27E-02
K00339			NADH dehydrogenase I subunit J [EC:1.6.5.3]	19	1	3.71	8.27E-04	5.84E-03
K00404			cb-type cytochrome c oxidase subunit I [EC:1.9.3.1]	0	53	NA	2.19E-21	1.96E-19
K04090		00191 Pyruvate/Oxoglutarate oxidoreductases	indolepyruvate ferredoxin oxidoreductase [EC:1.2.7.8]	13	0	NA	1.27E-03	8.72E-03
K00184			molybdopterin oxidoreductase, iron-sulfur binding subunit [EC:1.2.7.-]	10	0	NA	7.28E-03	3.78E-02
K02636		00195 Photosynthesis	cytochrome b6-f complex iron-sulfur subunit [EC:1.10.99.1]	2	10	-2.86	5.19E-03	2.78E-02
K02112			F-type H-transporting ATPase subunit beta [EC:3.6.3.14]	140	180	-0.90	2.60E-08	4.33E-07
K02115			F-type H-transporting ATPase subunit gamma [EC:3.6.3.14]	48	59	-0.84	3.02E-03	1.81E-02
K02705			photosystem II PsbC protein	352	108	1.17	6.79E-15	2.51E-13
K02703			photosystem II PsbA protein	560	154	1.32	1.32E-27	2.44E-25
K02704			photosystem II PsbB protein	235	61	1.41	2.19E-13	6.60E-12
K02689			photosystem I core protein Ia	86	20	1.57	2.26E-06	2.73E-05
K02699			photosystem I subunit XI	35	8	1.59	2.83E-03	1.71E-02
K02706			photosystem II PsbD protein	195	40	1.75	5.82E-15	2.21E-13
K02690			photosystem I core protein Ib	627	96	2.17	3.03E-59	1.40E-56
K02694			photosystem I subunit III	35	4	2.59	5.10E-05	4.97E-04
K02691			photosystem I subunit VII	15	1	3.37	3.92E-03	2.27E-02
K02708			photosystem II PsbF protein	12	0	NA	2.25E-03	1.39E-02
K02716			photosystem II PsbO protein	15	0	NA	7.36E-04	5.44E-03
K10713		00680 Methane metabolism	formaldehyde-activating enzyme [EC:4.3.-.-]	3	17	-3.04	9.13E-05	8.38E-04
K03520			carbon-monoxide dehydrogenase large subunit [EC:1.2.99.2]	34	4	2.55	8.45E-05	7.80E-04
K00122			formate dehydrogenase [EC:1.2.1.2]	26	2	3.16	1.38E-04	1.23E-03
K03518			carbon-monoxide dehydrogenase small subunit [EC:1.2.99.2]	15	0	NA	7.36E-04	5.44E-03
K00672			formylmethanofuran--tetrahydromethanopterin N-formyltransferase [EC:2.3.1.101]	0	8	NA	7.63E-04	5.45E-03
K00029		00710 Carbon fixation	malate dehydrogenase (oxaloacetate-decarboxylating)(NADP) [EC:1.1.1.40]	13	65	-2.86	1.56E-14	5.60E-13
K01601			ribulose-bisphosphate carboxylase large chain [EC:4.1.1.39]	439	140	1.11	4.72E-17	2.56E-15
K01602			ribulose-bisphosphate carboxylase small chain [EC:4.1.1.39]	90	25	1.31	2.34E-05	2.37E-04
K00174		00720 Reductive carboxylate cycle (CO2 fixation)	2-oxoglutarate ferredoxin oxidoreductase subunit alpha [EC:1.2.7.3]	12	0	NA	2.25E-03	1.39E-02
K01953		00910 Nitrogen metabolism	asparagine synthase (glutamine-hydrolysing) [EC:6.3.5.4]	9	29	-2.23	1.62E-05	1.69E-04
K00266			glutamate synthase (NADPH/NADH) small chain [EC:1.4.1.13] [1.4.1.14]	32	96	-2.12	6.12E-15	2.29E-13
K02198			cytochrome c-type biogenesis protein CcmF	10	22	-1.68	1.80E-03	1.18E-02
K00265			glutamate synthase (NADPH/NADH) large chain [EC:1.4.1.13] [1.4.1.14]	100	194	-1.49	3.63E-18	2.23E-16
K01915			glutamine synthetase [EC:6.3.1.2]	76	132	-1.34	6.09E-11	1.43E-09
K00605			aminomethyltransferase [EC:2.1.2.10]	66	13	1.81	5.08E-06	5.76E-05
K00459			2-nitropropane dioxygenase [EC:1.13.11.32]	18	2	2.63	4.98E-03	2.67E-02
K00264			glutamate synthase (NADPH) [EC:1.4.1.13]	19	2	2.71	2.99E-03	1.80E-02
K00260			glutamate dehydrogenase [EC:1.4.1.2]	0	8	NA	7.63E-04	5.45E-03
K00380		00920 Sulfur metabolism	sulfite reductase (NADPH) flavoprotein alpha-component [EC:1.8.1.2]	2	23	-4.06	1.21E-07	1.91E-06
K00390			phosphoadenosine phosphosulfate reductase [EC:1.8.4.8]	3	25	-3.60	1.33E-07	2.05E-06
K00640			serine O-acetyltransferase [EC:2.3.1.30]	3	16	-2.95	1.80E-04	1.55E-03
K00956			sulfate adenylyltransferase subunit 1 [EC:2.7.7.4]	4	18	-2.71	1.11E-04	1.00E-03
K00381			sulfite reductase (NADPH) hemoprotein beta-component [EC:1.8.1.2]	16	48	-2.12	5.38E-08	8.66E-07
K01738			cysteine synthase [EC:2.5.1.47]	29	44	-1.14	1.14E-03	7.85E-03
K00394			adenylylsulfate reductase, subunit A [EC:1.8.99.2]	35	5	2.27	1.48E-04	1.31E-03
K00647	01130 Lipid Metabolism	00061 Fatty acid biosynthesis	3-oxoacyl-[acyl-carrier-protein] synthase I [EC:2.3.1.41]	6	45	-3.45	2.46E-12	7.19E-11
K01716			3-hydroxydecanoyl-[acyl-carrier-protein] dehydratase [EC:4.2.1.60]	3	20	-3.28	1.24E-05	1.31E-04
K00648			3-oxoacyl-[acyl-carrier-protein] synthase III [EC:2.3.1.180] [EC:2.3.1.41]	13	31	-1.79	8.50E-05	7.83E-04
K01946			biotin carboxylase [EC:6.3.4.14]	30	7	1.56	6.81E-03	3.56E-02
K00666			fatty-acyl-CoA synthase [EC:6.2.1..]	29	4	2.32	5.51E-04	4.21E-03
K00208			enoyl-[acyl-carrier-protein] reductase I [EC:1.3.1.9]	23	3	2.40	2.10E-03	1.32E-02
K06445		00071 Fatty acid metabolism	acyl-CoA dehydrogenase [EC:1.3.99..]	13	254	-4.83	3.38E-81	4.68E-78
K05297			rubredoxin-NAD reductase [EC:1.18.1.1]	1	7	-3.35	9.63E-03	4.72E-02
K00632			acetyl-CoA acyltransferase [EC:2.3.1.16]	26	73	-2.03	4.23E-11	1.05E-09
K00022			3-hydroxyacyl-CoA dehydrogenase [EC:1.1.1.35]	47	112	-1.79	6.33E-14	2.04E-12
K01692			enoyl-CoA hydratase [EC:4.2.1.17]	86	19	1.64	1.17E-06	1.48E-05
K00626			acetyl-CoA C-acetyltransferase [EC:2.3.1.9]	88	15	2.01	1.44E-08	2.47E-07
K00128			aldehyde dehydrogenase (NAD) [EC:1.2.1.3]	128	15	2.55	1.55E-15	6.05E-14
K00496			alkane 1-monooxygenase [EC:1.14.15.3]	20	1	3.78	4.63E-04	3.60E-03
K00919		00100 Biosynthesis of steroids	4-diphosphocytidyl-2-C-methyl-D-erythritolinase [EC:2.7.1.148]	3	31	-3.91	1.10E-09	2.27E-08
K03526			4-hydroxy-3-methylbut-2-en-1-yl diphosphate synthase [EC:1.17.4.3]	6	29	-2.81	6.77E-07	9.29E-06
K01046		00561 Glycerolipid metabolism	triacylglycerol lipase [EC:3.1.1.3]	1	8	-3.54	4.38E-03	2.44E-02
K00981		00564 Glycerophospholipid metabolism	phosphatidate cytidylyltransferase [EC:2.7.7.41]	1	8	-3.54	4.38E-03	2.44E-02
K00995			CDP-diacylglycerol--glycerol-3-phosphate 3-phosphatidyltransferase [EC:2.7.8.5]	3	16	-2.95	1.80E-04	1.55E-03
K00057			glycerol-3-phosphate dehydrogenase (NAD(P)) [EC:1.1.1.94]	7	20	-2.05	6.01E-04	4.58E-03
K00432		00590 Arachidonic acid metabolism	glutathione peroxidase [EC:1.11.1.9]	11	0	NA	4.02E-03	2.31E-02
K01254			leukotriene-A4 hydrolase [EC:3.3.2.6]	0	6	NA	4.59E-03	2.48E-02
K01525	01140 Nucleotide Metabolism	00230 Purine metabolism	bis(5-nucleosyl)-tetraphosphatase (symmetrical) [EC:3.6.1.41]	1	9	-3.71	1.97E-03	1.25E-02
K01139			guanosine-3,5-bis(diphosphate) 3-pyrophosphohydrolase [EC:3.1.7.2]	3	19	-3.20	2.39E-05	2.42E-04
K01514			exopolyphosphatase [EC:3.6.1.11]	5	20	-2.54	1.11E-04	1.00E-03
K00759			adenine phosphoribosyltransferase [EC:2.4.2.7]	3	11	-2.41	5.31E-03	2.83E-02
K01939			adenylosuccinate synthase [EC:6.3.4.4]	10	34	-2.30	1.38E-06	1.72E-05
K01923			phosphoribosylaminoimidazole-succinocarboxamide synthase [EC:6.3.2.6]	5	17	-2.30	7.16E-04	5.32E-03
K01952			phosphoribosylformylglycinamidine synthase [EC:6.3.5.3]	21	34	-1.23	2.22E-03	1.39E-02
K00962			polynucleotide nucleotidyltransferase [EC:2.7.7.8]	148	50	1.03	6.08E-06	6.85E-05
K00525			ribonucleoside-diphosphate reductase alpha chain [EC:1.17.4.1]	75	22	1.23	2.55E-04	2.13E-03
K00087			xanthine dehydrogenase [EC:1.17.1.4]	59	4	3.34	1.14E-09	2.34E-08
K01516			nucleoside-triphosphatase [EC:3.6.1.15]	0	20	NA	1.61E-08	2.73E-07
K00945		00240 Pyrimidine metabolism	cytidylateinase [EC:2.7.4.14]	8	43	-2.97	2.81E-10	6.09E-09
K09963			undylateinase [EC:2.7.4.22]	23	35	-1.14	3.04E-03	1.82E-02
K00818	01150 Amino Acid Metabolism	00220 Urea cycle and metabolism of amino groups	acetylornithine aminotransferase [EC:2.6.1.11]	4	27	-3.29	2.17E-07	3.18E-06
K00931			glutamate 5-kinase [EC:2.7.2.11]	4	13	-2.24	4.81E-03	2.59E-02
K00611			ornithine carbamoyltransferase [EC:2.1.3.3]	10	32	-2.22	4.48E-06	5.15E-05
K01940			argininosuccinate synthase [EC:6.3.4.5]	13	29	-1.70	2.42E-04	2.04E-03
K01585		00251 Glutamate metabolism	arginine decarboxylase [EC:4.1.1.19]	2	15	-3.45	7.42E-05	6.92E-04
K01580			glutamate decarboxylase [EC:4.1.1.15]	0	8	NA	7.63E-04	5.45E-03
K00609		00252 Alanine and aspartate metabolism	aspartate carbamoyltransferase catalytic subunit [EC:2.1.3.2]	4	12	-2.12	8.64E-03	4.33E-02
K00836		00260 Glycine, serine and threonine metabolism	diaminobutyrate-2-oxoglutarate transaminase [EC:2.6.1.76]	1	33	-5.58	2.83E-12	8.16E-11

Supplemental Table S6 (cont.): Pairwise tests of Functional Annotations Between Controls (Pooled) and HMWDOM Treatments 27 Hours Post Addition *

KO	Functional category	Pathway	ORF Annotation	Control	DOM27	ln(Fold change)	p-value	q-value
K01079			phosphoserine phosphatase [EC:3.1.3.3]	1	7	-3.35	9.63E-03	4.72E-02
K01754			threonine dehydratase [EC:4.3.1.19]	6	37	-3.16	1.28E-09	2.60E-08
K01878			glycyl-tRNA synthetase alpha chain [EC:6.1.1.14]	3	11	-2.41	5.31E-03	2.83E-02
K00872			homoserinekinase [EC:2.7.1.39]	5	14	-2.02	4.45E-03	2.47E-02
K00133			aspartate-semialdehyde dehydrogenase [EC:1.2.1.11]	12	23	-1.48	3.22E-03	1.91E-02
K01733			threonine synthase [EC:4.2.3.1]	16	30	-1.45	1.33E-03	8.97E-03
K00928			aspartatekinase [EC:2.7.2.4]	17	30	-1.36	1.63E-03	1.07E-02
K00315			dimethylglycine dehydrogenase [EC:1.5.99.2]	38	6	2.12	1.58E-04	1.39E-03
K02204			homoserinekinase type II [EC:2.7.1.39]	0	7	NA	1.87E-03	1.19E-02
K06718			L-2,4-diaminobutyric acid acetyltransferase [EC:2.3.1.178]	0	10	NA	1.27E-04	1.14E-03
K06720			L-ectoine synthase [EC:4.2.1.108]	0	16	NA	5.82E-07	8.02E-06
K00828			serine--glyoxylate transaminase [EC:2.6.1.45]	11	0	NA	4.02E-03	2.31E-02
K01740		00271 Methionine metabolism	O-acetylhomoserine (thiol)-lyase [EC:2.5.1.49]	119	24	1.77	1.09E-09	2.27E-08
K10764			O-succinylhomoserine sulfhydrylase [EC:2.5.1.-]	0	7	NA	1.87E-03	1.19E-02
K00053		00290 Valine, leucine and isoleucine biosynthesis	ketol-acid reductoisomerase [EC:1.1.1.86]	36	127	-2.36	4.92E-22	4.71E-20
K00052			3-isopropylmalate dehydrogenase [EC:1.1.1.85]	23	39	-1.30	6.44E-04	4.87E-03
K01687			dihydroxy-acid dehydratase [EC:4.2.1.9]	35	45	-0.90	6.03E-03	3.18E-02
K00003		00300 Lysine biosynthesis	homoserine dehydrogenase [EC:1.1.1.3]	10	49	-2.83	3.24E-11	8.62E-10
K01439			succinyl-diaminopimelate desuccinylase [EC:3.5.1.18]	8	19	-1.79	2.62E-03	1.60E-02
K00674			2,3,4,5-tetrahydropyridine-2-carboxylate N-succinyltransferase [EC:2.3.1.117]	10	23	-1.74	1.09E-03	7.53E-03
K00290			saccharopine dehydrogenase (NAD, L-lysine forming) [EC:1.5.1.7]	0	7	NA	1.87E-03	1.19E-02
K06447		00330 Arginine and proline metabolism	succinylglutamic semialdehyde dehydrogenase [EC:1.2.1.71]	1	7	-3.35	9.63E-03	4.72E-02
K00294			1-pyrroline-5-carboxylate dehydrogenase [EC:1.5.1.12]	4	17	-2.63	2.21E-04	1.88E-03
K01693		00340 Histidine metabolism	imidazoleglycerol-phosphate dehydratase [EC:4.2.1.19]	16	1	3.46	2.29E-03	1.42E-02
K01636		00350 Tyrosine metabolism	hexulose-6-phosphate synthase [EC:4.1.2.-] (EC:4.1.2.43)	0	46	NA	1.17E-18	7.40E-17
K04093		00400 Phenylalanine, tyrosine and tryptophan biosynthesis	chorismate mutase [EC:5.4.99.5]	1	10	-3.86	8.78E-04	6.14E-03
K01736			chorismate synthase [EC:4.2.3.5]	10	24	-1.80	6.53E-04	4.93E-03
K01657			anthranilate synthase component I [EC:4.1.3.27]	7	16	-1.73	9.16E-03	4.56E-02
K01626			3-deoxy-7-phosphoheptulonate synthase [EC:2.5.1.54]	28	38	-0.98	7.99E-03	4.11E-02
K01696			tryptophan synthase beta chain [EC:4.2.1.20]	33	8	1.51	6.19E-03	3.25E-02
K04092			chorismate mutase [EC:5.4.99.5]	0	25	NA	1.81E-10	3.97E-09
K00210			prephenate dehydrogenase [EC:1.3.1.12]	0	16	NA	5.82E-07	8.02E-06
K03119	01160 Metabolism of Other Amino Acids	00430 Taurine and hypotaurine metabolism	taurine dioxygenase [EC:1.14.11.17]	27	0	NA	1.01E-06	1.38E-05
K01763		00450 Selenoamino acid metabolism	selenocysteine lyase [EC:4.4.1.16]	14	26	-1.43	3.21E-03	1.91E-02
K02517	01170 Glycan Biosynthesis and Metabolism	00540 Lipopolysaccharide biosynthesis	lipid A biosynthesis lauroyl acyltransferase [EC:2.3.1.-]	1	11	-4.00	3.89E-04	3.05E-03
K02535			UDP-3-O-[3-hydroxymyristoyl] N-acetylglucosamine deacetylase [EC:3.5.1.-]	20	44	-1.68	9.01E-06	9.64E-05
K00748			lipid-A-disaccharide synthase [EC:2.4.1.182]	0	13	NA	8.59E-06	9.30E-05
K01661	01190 Metabolism of Cofactors and Vitamins	00130 Ubiquinone biosynthesis	naphthoate synthase [EC:4.1.3.36]	12	0	NA	2.25E-03	1.39E-02
K01934		00670 One carbon pool by folate	5-formyltetrahydrofolate cyclo-ligase [EC:6.3.3.2]	2	9	-2.71	1.00E-02	4.86E-02
K00794		00740 Riboflavin metabolism	riboflavin synthase beta chain [EC:2.5.1.-]	2	17	-3.63	1.54E-05	1.62E-04
K03473		00750 Vitamin B6 metabolism	erythronate-4-phosphate dehydrogenase [EC:1.1.1.290]	1	7	-3.35	9.63E-03	4.72E-02
K00325		00760 Nicotinate and nicotinamide metabolism	NAD(P) transhydrogenase subunit beta [EC:1.6.1.2]	29	120	-2.59	3.45E-23	4.55E-21
K00324			NAD(P) transhydrogenase subunit alpha [EC:1.6.1.2]	27	84	-2.18	9.22E-14	2.90E-12
K01423		00780 Biotin metabolism	peptidase, M28 (aminopeptidase S) family (EC:3.4.-.-)	12	28	-1.76	2.76E-04	2.28E-03
K01927		00790 Folate biosynthesis	dihydrofolate synthase [EC:6.3.2.12]	2	15	-3.45	7.42E-05	6.92E-04
K01077			alkaline phosphatase [EC:3.1.3.1]	5	26	-2.92	1.13E-06	1.48E-05
K03639			molybdenum cofactor biosynthesis protein A	4	15	-2.45	1.48E-03	9.81E-03
K01529			RecG-like helicase	9	24	-1.95	2.65E-04	2.20E-03
K01495			GTP cyclohydrolase I [EC:3.5.4.16]	22	4	1.92	8.48E-03	4.33E-02
K01737			6-pyruvoyl tetrahydrobiopterin synthase [EC:4.2.3.12]	42	3	3.27	3.85E-07	5.50E-06
K03403		00860 Porphyrin and chlorophyll metabolism	magnesium chelatase [EC:6.6.1.1]	22	4	1.92	8.48E-03	4.33E-02
K02496			uroporphyrin-III C-methyltransferase [EC:2.1.1.107]	0	13	NA	8.59E-06	9.30E-05
K01835	01195 Biosynthesis of Secondary Metabolites	00521 Streptomycin biosynthesis	phosphoglucomutase [EC:5.4.2.2]	1	18	-4.71	1.13E-06	1.48E-05
K01904		00940 Phenylpropanoid biosynthesis	4-coumarate--CoA ligase [EC:6.2.1.12]	0	7	NA	1.87E-03	1.19E-02
K00257	01196 Xenobiotics Biodegradation and Metabolism	00281 Geraniol degradation	acyl-CoA dehydrogenase	153	70	0.59	4.13E-03	2.36E-02
K10680		00633 Trinitrotoluene degradation	N-ethylmaleimide reductase [EC:1.-.-.-]	1	9	-3.71	1.97E-03	1.25E-02
K03379		00930 Caprolactam degradation	cyclohexanone monooxygenase [EC:1.14.13.22]	23	3	2.40	2.10E-03	1.32E-02
K00001		00980 Metabolism of xenobiotics by cytochrome P450	alcohol dehydrogenase [EC:1.1.1.1]	43	92	-1.64	1.74E-10	3.86E-09
K03088	01210 Transcription	03020 RNA polymerase	RNA polymerase sigma-70 factor, ECF subfamily	19	43	-1.72	6.70E-06	7.51E-05
K03040			DNA-directed RNA polymerase subunit alpha [EC:2.7.7.6]	91	182	-1.54	8.88E-18	5.13E-16
K03043			DNA-directed RNA polymerase subunit beta [EC:2.7.7.6]	213	238	-0.70	3.16E-07	4.53E-06
K03046			DNA-directed RNA polymerase subunit beta [EC:2.7.7.6]	312	297	-0.47	6.96E-05	6.63E-04
K03042			DNA-directed RNA polymerase subunit A [EC:2.7.7.6]	0	6	NA	4.59E-03	2.48E-02
K03628		03028 Other transcription related proteins	transcription termination factor Rho	14	39	-2.02	1.75E-06	2.14E-05
K02601			transcriptional antiterminator NusG	47	83	-1.36	1.38E-07	2.12E-06
K02500			N utilization substance protein A	48	77	-1.22	3.86E-06	4.57E-05
K03624			transcription elongation factor GreA	0	8	NA	7.63E-04	5.45E-03
K02502	01220 Translation	00970 Aminoacyl-tRNA biosynthesis	ATP phosphorobosyltransferase regulatory subunit	2	11	-3.00	2.69E-03	1.64E-02
K01875			seryl-tRNA synthetase [EC:6.1.1.11]	4	19	-2.79	5.54E-05	5.33E-04
K01883			cysteinyl-tRNA synthetase [EC:6.1.1.16]	7	25	-2.38	3.25E-05	3.24E-04
K01868			threonyl-tRNA synthetase [EC:6.1.1.3]	17	60	-2.36	4.40E-11	1.07E-09
K01886			glutaminyl-tRNA synthetase [EC:6.1.1.18]	4	12	-2.12	8.64E-03	4.33E-02
K01867			tryptophanyl-tRNA synthetase [EC:6.1.1.2]	6	17	-2.04	2.12E-03	1.33E-02
K01869			leucyl-tRNA synthetase [EC:6.1.1.4]	15	42	-2.02	5.58E-07	7.89E-06
K01873			valyl-tRNA synthetase [EC:6.1.1.9]	14	33	-1.78	7.14E-05	6.75E-04
K01887			arginyl-tRNA synthetase [EC:6.1.1.19]	7	16	-1.73	9.16E-03	4.56E-02
K01892			histidyl-tRNA synthetase [EC:6.1.1.21]	9	20	-1.69	3.71E-03	2.18E-02
K01881			prolyl-tRNA synthetase [EC:6.1.1.15]	27	6	1.63	7.53E-03	3.87E-02
K02919		03010 Ribosome	large subunit ribosomal protein L36	8	37	-2.75	1.63E-08	2.75E-07
K02911			large subunit ribosomal protein L32	21	97	-2.75	3.10E-20	2.60E-18
K02897			large subunit ribosomal protein L25	24	104	-2.65	1.16E-20	1.01E-18
K02939			large subunit ribosomal protein L9	22	94	-2.63	8.44E-19	5.43E-17
K02954			small subunit ribosomal protein S14	37	133	-2.38	4.41E-23	5.55E-21
K02931			large subunit ribosomal protein L5	75	267	-2.37	5.86E-45	1.80E-42
K02956			small subunit ribosomal protein S15	14	47	-2.29	1.05E-08	2.00E-07
K02914			large subunit ribosomal protein L34	3	10	-2.28	1.02E-02	4.89E-02
K02874			large subunit ribosomal protein L14	30	95	-2.20	1.39E-15	6.05E-14
K02990			small subunit ribosomal protein S6	16	47	-2.09	9.14E-08	1.45E-06
K02899			large subunit ribosomal protein L27	19	54	-2.05	1.23E-08	2.33E-07
K02888			large subunit ribosomal protein L21	26	73	-2.03	4.23E-11	1.05E-09
K02867			large subunit ribosomal protein L11	30	82	-1.99	5.36E-12	1.51E-10
K02879			large subunit ribosomal protein L17	30	74	-1.84	6.06E-10	1.29E-08
K02945			small subunit ribosomal protein S11	98	223	-1.72	3.40E-25	5.53E-23
K02933			large subunit ribosomal protein L6	44	100	-1.72	5.18E-12	1.48E-10
K02906			large subunit ribosomal protein L3	71	158	-1.69	8.86E-18	5.13E-16
K02876			large subunit ribosomal protein L15	38	84	-1.68	4.39E-10	9.42E-09
K02916			large subunit ribosomal protein L35	13	28	-1.65	6.31E-04	4.79E-03
K02994			small subunit ribosomal protein S8	51	107	-1.61	9.91E-12	2.75E-10
K02961			small subunit ribosomal protein S17	15	30	-1.54	6.56E-04	4.94E-03
K02887			large subunit ribosomal protein L20	24	47	-1.51	1.73E-05	1.80E-04
K02986			small subunit ribosomal protein S4	87	162	-1.44	1.30E-14	4.73E-13
K02992			small subunit ribosomal protein S7	53	92	-1.33	4.00E-08	6.56E-07
K02890			large subunit ribosomal protein L22	29	48	-1.27	1.68E-04	1.47E-03
K02904			large subunit ribosomal protein L29	17	28	-1.26	5.65E-03	2.99E-02
K02878			large subunit ribosomal protein L16	49	79	-1.23	2.11E-06	2.57E-05
K02988			small subunit ribosomal protein S5	66	106	-1.22	3.77E-08	6.21E-07
K02952			small subunit ribosomal protein S13	96	152	-1.20	8.49E-11	1.90E-09
K02965			small subunit ribosomal protein S19	35	55	-1.19	1.42E-04	1.26E-03
K02871			large subunit ribosomal protein L13	33	51	-1.17	3.20E-04	2.58E-03
K02935			large subunit ribosomal protein L7/L12	154	220	-1.05	2.08E-12	6.14E-11
K02926			large subunit ribosomal protein L4	64	90	-1.03	1.14E-05	1.21E-04
K02967			small subunit ribosomal protein S2	60	84	-1.02	2.69E-05	2.71E-04
K02886			large subunit ribosomal protein L2	95	132	-1.01	1.49E-07	2.26E-06
K02863			large subunit ribosomal protein L1	94	130	-1.01	2.53E-07	3.68E-06
K02895			large subunit ribosomal protein L24	47	60	-0.89	1.55E-03	1.02E-02
K02982			small subunit ribosomal protein S3	99	118	-0.79	5.60E-05	5.37E-04
K02950			small subunit ribosomal protein S12	45	53	-0.77	9.78E-03	4.78E-02
K02948			small subunit ribosomal protein S11	69	79	-0.73	2.46E-03	1.51E-02
K02528		03014 Other translation proteins	dimethyladenosine transferase [EC:2.1.1.-]	1	10	-3.86	8.78E-04	6.14E-03
K00791			tRNA delta(2)-isopentenylpyrophosphate transferase [EC:2.5.1.8]	3	25	-3.60	1.33E-07	2.05E-06
K06176			tRNA pseudouridine synthase D [EC:5.4.99.12]	1	8	-3.54	4.38E-03	2.44E-02
K00554			tRNA (guanine-N1-)-methyltransferase [EC:2.1.1.31]	28	208	-3.43	8.49E-52	3.36E-49
K02860			16S rRNA processing protein RimM	20	105	-2.93	3.05E-23	4.23E-21
K06158			ATP-binding cassette, sub-family F, member 3	4	19	-2.79	5.54E-05	5.33E-04
K03500			ribosomal RNA small subunit methyltransferase B [EC:2.1.1.-]	2	9	-2.71	1.00E-02	4.86E-02
K06178			ribosomal large subunit pseudouridine synthase B [EC:5.4.99.12]	4	17	-2.63	2.21E-04	1.88E-03
K05540			RNA-dihydrouridine synthase B [EC:1.-.-.-]	12	50	-2.60	2.60E-10	5.66E-09
K06168			bifunctional enzyme involved in thiolation and methylation of tRNA	8	30	-2.45	2.64E-06	3.16E-05
K00970			poly(A) polymerase [EC:2.7.7.19]	4	15	-2.45	1.48E-03	9.81E-03
K02834			ribosome-binding factor A	3	10	-2.28	1.02E-02	4.89E-02
K00773			queuine tRNA-ribosyltransferase [EC:2.4.2.29]	4	12	-2.12	8.64E-03	4.33E-02
K08300			ribonuclease E [EC:3.1.4.-]	14	41	-2.09	5.77E-07	8.02E-06
K03536			ribonuclease P protein component [EC:3.1.26.5]	5	14	-2.02	4.45E-03	2.47E-02
K03685			ribonuclease III [EC:3.1.26.3]	5	13	-1.92	8.04E-03	4.12E-02
K01056			peptidyl-tRNA hydrolase, PTH1 family [EC:3.1.1.29]	8	18	-1.71	4.39E-03	2.44E-02

Supplemental Table S6 (cont.): Pairwise tests of Functional Annotations Between Controls (Pooled) and HMWDOM Treatments 27 Hours Post Addition *

KO	Functional category	Pathway	ORF Annotation	Control	DOM27	ln(Fold change)	p-value	q-value
K01975			2-5 RNA ligase [EC:6.5.1.-]	0	6	NA	4.59E-03	2.48E-02
K07576			metallo-beta-lactamase family protein	0	8	NA	7.63E-04	5.45E-03
K05844			ribosomal protein S6 modification protein	0	7	NA	1.87E-03	1.19E-02
K03215			RNA methyltransferase, TrmA family [EC:2.1.1.-]	0	13	NA	8.59E-06	9.30E-05
K03071	01230 Folding, Sorting and Degradation	03060 Protein export	preprotein translocase SecB subunit	2	12	-3.12	1.40E-03	9.31E-03
K03076			preprotein translocase SecY subunit	75	351	-2.77	7.39E-70	6.83E-67
K02453		03090 Type II secretion system	general secretion pathway protein D	2	13	-3.24	3.49E-04	2.79E-03
K02454			general secretion pathway protein E	13	25	-1.48	2.50E-03	1.53E-02
K02455			general secretion pathway protein F	0	6	NA	4.59E-03	2.48E-02
K02456			general secretion pathway protein G	10	0	NA	7.28E-03	3.78E-02
K02458			general secretion pathway protein I	0	6	NA	4.59E-03	2.48E-02
K02656			type IV pilus assembly protein PilF	0	6	NA	4.59E-03	2.48E-02
K02663			type IV pilus assembly protein PilN	0	6	NA	4.59E-03	2.48E-02
K06891		03100 Protein folding and associated processing	ATP-dependent Clp protease adaptor protein ClpS	1	12	-4.12	1.71E-04	1.49E-03
K00982			glutamate-ammonia-ligase adenylyltransferase [EC:2.7.7.42]	1	12	-4.12	1.71E-04	1.49E-03
K03770			peptidyl-prolyl cis-trans isomerase D [EC:5.2.1.8]	7	37	-2.94	7.30E-09	1.41E-07
K07399			cytochrome c biogenesis protein	2	9	-2.71	1.00E-02	4.86E-02
K03545			trigger factor	37	71	-1.48	2.04E-07	3.07E-06
K04088			membrane protease subunit HflK [EC:3.4.-.-]	36	54	-1.12	3.41E-04	2.74E-03
K03386			peroxiredoxin (alkyl hydroperoxide reductase subunit C) [EC:1.11.1.15]	43	62	-1.07	2.04E-04	1.75E-03
K07400			thioredoxin-like protein	19	3	2.12	8.65E-03	4.33E-02
K01802			peptidylprolyl isomerase [EC:5.2.1.8]	34	4	2.55	8.45E-05	7.80E-04
K03775			FKBP-type peptidyl-prolyl cis-trans isomerase SlyD [EC:5.2.1.8]	0	22	NA	2.67E-09	5.21E-08
K02316	01240 Replication and Repair	03030 DNA replication	DNA primase [EC:2.7.7.-]	5	19	-2.46	2.07E-04	1.77E-03
K03469			ribonuclease HI [EC:3.1.26.4]	19	2	2.71	2.99E-03	1.80E-02
K07497		03034 Other replication, recombination and repair proteins	putative transposase	1	9	-3.71	1.97E-03	1.25E-02
K03655			ATP-dependent DNA helicase RecG [EC:3.6.1.-]	1	8	-3.54	4.38E-03	2.44E-02
K04096			DNA processing protein	1	7	-3.35	9.63E-03	4.72E-02
K03732			ATP-dependent RNA helicase RhlB [EC:3.6.1.-]	2	13	-3.24	3.49E-04	2.79E-03
K03722			ATP-dependent DNA helicase DinG [EC:3.6.1.-]	3	13	-2.65	1.39E-03	9.26E-03
K00986			RNA-directed DNA polymerase [EC:2.7.7.49]	3	12	-2.54	2.73E-03	1.65E-02
K03580			ATP-dependent helicase HepA [EC:3.6.1.-]	0	7	NA	1.87E-03	1.19E-02
K05591			ATP-independent RNA helicase DbpA	0	6	NA	4.59E-03	2.48E-02
K07493			putative transposase	0	26	NA	7.37E-11	1.71E-09
K07486			transposase	0	22	NA	2.67E-09	5.21E-08
K10773		03038 Base excision repair	endonuclease III [EC:4.2.99.18]	1	9	-3.71	1.97E-03	1.25E-02
K03648			uracil-DNA glycosylase [EC:3.2.2.-]	0	9	NA	3.11E-04	2.53E-03
K02038	01310 Membrane Transport	02010 ABC transporters	phosphate transport system permease protein	3	74	-5.16	2.21E-25	3.83E-23
K02037			phosphate transport system permease protein	8	111	-4.33	6.49E-34	1.50E-31
K02036			phosphate transport system ATP-binding protein [EC:3.6.3.27]	5	55	-4.00	1.54E-16	8.23E-15
K07323			putative toluene tolerance protein	1	11	-4.00	3.89E-04	3.05E-03
K02040			phosphate transport system substrate-binding protein	32	305	-3.79	2.98E-82	8.24E-79
K06148			ATP-binding cassette, subfamily C, bacterial	1	7	-3.35	9.63E-03	4.72E-02
K02066			putative ABC transport system permease protein	7	35	-2.86	2.65E-08	4.39E-07
K02065			putative ABC transport system ATP-binding protein	5	23	-2.74	9.48E-06	1.01E-04
K02055			putative spermidine/putrescine transport system substrate-binding protein	111	44	0.80	1.79E-03	1.17E-02
K01999			branched-chain amino acid transport system substrate-binding protein	158	49	1.15	2.68E-07	3.87E-06
K02035			peptide/nickel transport system substrate-binding protein	127	38	1.20	2.25E-06	2.72E-05
K02051			sulfonate/nitrate/taurine transport system substrate-binding protein	52	14	1.35	1.01E-03	7.04E-03
K02002			glycine betaine/proline transport system substrate-binding protein	86	21	1.50	4.34E-06	5.03E-05
K09969			general L-amino acid transport system substrate-binding protein	141	20	2.28	3.58E-15	1.38E-13
K01998			branched-chain amino acid transport system permease protein	19	1	3.71	8.27E-04	5.84E-03
K05559		02052 Other ion-coupled transporters	multicomponent:H+ antiporter subunit A	1	8	-3.54	4.38E-03	2.44E-02
K07239			heavy-metal exporter, HME family	3	21	-3.35	3.03E-06	3.62E-05
K03327			multidrug resistance protein, MATE family	2	9	-2.71	1.00E-02	4.86E-02
K03306			inorganic phosphate transporter, PiT family	6	19	-2.20	7.05E-04	5.29E-03
K07301			ecm27	5	15	-2.12	2.44E-03	1.50E-02
K03296			hydrophobic/amphiphilic exporter-1 (mainly G- bacteria), HAE1 family	16	25	-1.18	1.05E-02	4.99E-02
K03307			solute:Na symporter, SSS family	296	68	1.58	2.57E-19	1.88E-17
K03320			ammonium transporter, Amt family	718	157	1.65	7.71E-48	2.67E-45
K03325			arsenite transporter, ACR3 family	0	6	NA	4.59E-03	2.48E-02
K02168			high-affinity choline transport protein	0	44	NA	7.08E-16	4.26E-14
K03315			Na:H antiporter, NhaC family	0	8	NA	7.63E-04	5.45E-03
K08484		02060 Phosphotransferase system (PTS)	phosphotransferase system, enzyme I, PtsP [EC:2.7.3.9]	1	8	-3.54	4.38E-03	2.44E-02
K03286		02070 Pores ion channels	OmpA-OmpF porin, OOP family	5	94	-4.77	1.21E-30	2.39E-28
K03442			small conductance mechanosensitive ion channel, MscS family	6	32	-2.95	4.81E-08	7.85E-07
K03287			outer membrane factor, OMF family	5	13	-1.92	8.04E-03	4.12E-02
K02014			iron complex outermembrane recepter protein	184	318	-1.33	2.16E-24	3.14E-22
K08720			outer membrane protein OmpU	51	13	1.43	8.00E-04	5.69E-03
K05802			potassium efflux system protein#A	0	6	NA	4.59E-03	2.48E-02
K03611		02080 Electron transfer carriers	disulfide bond formation protein DsbB	0	7	NA	1.87E-03	1.19E-02
K06189		02082 Other transporters	magnesium and cobalt transporter	4	43	-3.97	4.07E-13	1.21E-11
K03811			nicotinamide mononucleotide transporter	0	6	NA	4.59E-03	2.48E-02
K07657	01320 Signal Transduction	02020 Two-component system	two-component system, OmpR family, phosphate regulon response regulator PhoB	2	91	-6.05	5.02E-33	1.07E-30
K07678			two-component system, NarL family, sensor histidinekinase BarA [EC:2.7.13.3]	1	14	-4.35	3.26E-05	3.24E-04
K03413			two-component system, chemotaxis family, response regulator CheY	2	26	-4.24	1.03E-08	1.96E-07
K03407			two-component system, chemotaxis family, sensorinase CheA [EC:2.7.13.3]	6	64	-3.95	6.81E-19	4.49E-17
K07814			putative two-component system response regulator	1	10	-3.86	8.78E-04	6.14E-03
K03563			carbon storage regulator	2	18	-3.71	6.93E-06	7.71E-05
K07708			two-component system, NtrC family, nitrogen regulation sensor histidinekinase GlnL [EC:2.7.13.3]	1	7	-3.35	9.63E-03	4.72E-02
K07715			two-component system, NtrC family, response regulator YhlA	7	46	-3.25	6.04E-12	1.69E-10
K02488			two-component system, PleD related family, response regulator	4	25	-3.18	8.42E-07	1.15E-05
K07712			two-component system, NtrC family, nitrogen regulation response regulator GlnG	7	26	-2.43	1.82E-05	1.89E-04
K07795			putative tricarboxylic transport membrane protein	56	6	2.68	8.21E-08	1.31E-06
K03412			protein-glutamate methylesterase, two-component system, chemotaxis family, response regulator CheB [EC:3.1.1.61]	0	27	NA	3.00E-11	8.07E-10
K02106			short-chain fatty acids transporter	0	15	NA	1.43E-06	1.77E-05
K03415			two-component system, chemotaxis family, response regulator CheV	0	13	NA	8.59E-06	9.30E-05
K07689			two-component system, NarL family, invasion response regulator UvrY	0	7	NA	1.87E-03	1.19E-02
K07675			two-component system, NarL family, sensor histidinekinase UhpB [EC:2.7.13.3]	0	6	NA	4.59E-03	2.48E-02
K07711			two-component system, NtrC family, sensor histidinekinase YfhK [EC:2.7.13.3]	0	27	NA	3.00E-11	8.07E-10
K07773			two-component system, OmpR family, aerobic respiration control protein ArcA	0	17	NA	1.37E-07	3.46E-06
K07636			two-component system, OmpR family, phosphate regulon sensor histidinekinase PhoR [EC:2.7.13.3]	0	57	NA	6.03E-23	7.26E-21
K07662			two-component system, OmpR family, response regulator CpxR	0	7	NA	1.87E-03	1.19E-02
K03283		04010 MAPK signaling pathway	heat shock 70kDa protein 1/8	13	0	NA	1.27E-03	8.72E-03
K03408	01410 Cell Motility	02030 Bacterial chemotaxis	purine-binding chemotaxis protein CheW	1	37	-5.75	8.70E-14	2.77E-12
K00575			chemotaxis protein methyltransferase CheR [EC:2.1.1.80]	2	41	-4.90	3.43E-14	1.22E-12
K03414			chemotaxis protein CheZ	2	13	-3.24	3.49E-04	2.79E-03
K03410			chemotaxis protein CheC	2	9	-2.71	1.00E-02	4.86E-02
K03406			methyl-accepting chemotaxis protein	18	77	-2.64	1.21E-15	6.05E-14
K02399		02040 Flagellar assembly	flagella synthesis protein FlgN	1	11	-4.00	3.89E-04	3.05E-03
K02391			flagellar basal-body rod protein FlgF	1	10	-3.86	8.78E-04	6.14E-03
K02411			flagellar assembly protein FliH	1	8	-3.54	4.38E-03	2.44E-02
K02407			flagellar hook-associated protein 2	10	76	-3.46	7.54E-20	6.14E-18
K02393			flagellar L-ring protein precursor FlgH	4	22	-3.00	6.77E-06	7.56E-05
K06603			flagellar protein FlaG	2	9	-2.71	1.00E-02	4.86E-02
K02404			flagellar biosynthesis protein FlhF	5	21	-2.61	6.00E-05	5.73E-04
K02409			flagellar M-ring protein FliF	10	34	-2.30	1.38E-06	1.72E-05
K02394			flagellar P-ring protein precursor FlgI	14	41	-2.09	5.77E-07	8.02E-06
K02416			flagellar motor switch protein FliM	9	20	-1.69	3.71E-03	2.18E-02
K04562			flagellar biosynthesis protein FlhG	14	29	-1.59	5.07E-04	3.94E-03
K02556			chemotaxis protein MotA	13	24	-1.42	3.92E-03	2.27E-02
K02406			flagellin	260	298	-0.74	2.07E-09	4.10E-08
K02408			flagellar hook-basal body complex protein FliE	18	2	2.63	4.98E-03	2.67E-02
K02396			flagellar hook-associated protein 1 FlgK	0	14	NA	3.50E-06	4.16E-05
K02395			flagellar protein FlgJ	0	11	NA	5.17E-05	5.03E-04
K02427	01420 Cell Growth and Death	04410 Cell division	cell division protein methyltransferase FtsJ [EC:2.1.1.-]	1	11	-4.00	3.89E-04	3.05E-03
K03570			rod shape-determining protein MreC	1	8	-3.54	4.38E-03	2.44E-02
K03528			cell division protein ZipA	3	12	-2.54	2.73E-03	1.65E-02
K03496			chromosome partitioning protein	10	28	-2.02	4.70E-05	4.63E-04
K03798			cell division protease FtsH [EC:3.4.24.-]	209	97	0.57	1.32E-03	8.97E-03
K06252	01430 Cell Communication	04510 Focal adhesion	tenascin	16	0	NA	3.91E-04	3.05E-03

Supplemental Table S6 (cont.): Pairwise tests of Functional Annotations Between HMWDOM Treatments 12 Hours Post Addition and HMWDOM Treatments 27 Hours Post Addition *

KO	Functional category	Pathway	ORF Annotation	DOM12	DOM27	ln(Fold change)	p-value	q-value
K00616	01110 Carbohydrate Metabolism	00030 Pentose phosphate pathway	transaldolase [EC:2.2.1.2]	2	39	-2.85	5.88E-04	3.43E-02
K08093		00040 Pentose and glucuronate interconversions	3-hexulose-6-phosphate synthase [EC:4.1.2.-]	10	79	-1.54	4.52E-04	2.71E-02
K08094			6-phospho-3-hexuloisomerase [EC:5.-.-.-]	1	93	-5.10	7.04E-12	7.58E-09
K01199		00500 Starch and sucrose metabolism	glucan endo-1,3-beta-D-glucosidase [EC:3.2.1.39]	8	0	NA	2.78E-05	3.71E-03
K00790		00530 Aminosugars metabolism	UDP-N-acetylglucosamine 1-carboxyvinyltransferase [EC:2.5.1.7]	17	12	1.94	4.37E-04	2.70E-02
K01007		00620 Pyruvate metabolism	pyruvate,water dikinase [EC:2.7.9.2]	104	173	0.70	1.33E-04	1.15E-02
K00412	01120 Energy Metabolism	00190 Oxidative phosphorylation	ubiquinol-cytochrome c reductase cytochrome b subunit [EC:1.10.2.2]	38	39	1.40	2.82E-05	3.71E-03
K02116		00195 Photosynthesis	ATP synthase protein I	6	0	NA	3.83E-04	2.50E-02
K02111			F-type H-transporting ATPase subunit alpha [EC:3.6.3.14]	85	118	0.97	5.05E-06	1.21E-03
K02112			F-type H-transporting ATPase subunit beta [EC:3.6.3.14]	102	180	0.62	7.34E-04	3.68E-02
K02690			photosystem I core protein Ib	75	96	1.08	2.33E-06	7.18E-04
K01601		00710 Carbon fixation	ribulose-bisphosphate carboxylase large chain [EC:4.1.1.39]	14	140	-1.88	4.79E-08	2.58E-05
K00264		00910 Nitrogen metabolism	glutamate synthase (NADPH) [EC:1.4.1.13]	9	2	3.61	2.36E-04	1.89E-02
K00128	01130 Lipid Metabolism	00071 Fatty acid metabolism	aldehyde dehydrogenase (NAD) [EC:1.2.1.3]	18	15	1.70	1.14E-03	3.84E-02
K00951	01140 Nucleotide Metabolism	00230 Purine metabolism	GTP pyrophosphokinase [EC:2.7.6.5]	12	7	2.22	1.01E-03	3.84E-02
K01945			phosphoribosylamine--glycine ligase [EC:6.3.4.13]	14	6	2.66	7.30E-05	8.74E-03
K00087			xanthine dehydrogenase [EC:1.17.1.4]	10	4	2.76	6.60E-04	3.64E-02
K00618		00220 Urea cycle and metabolism of amino groups	acetylornithine aminotransferase [EC:2.6.1.11]	0	27	NA	3.02E-04	2.24E-02
K00836		00260 Glycine, serine and threonine metabolism	diaminobutyrate-2-oxoglutarate transaminase [EC:2.6.1.76]	1	33	-3.61	6.55E-04	3.64E-02
K00282			glycine dehydrogenase subunit 1 [EC:1.4.4.2]	20	14	1.95	1.14E-04	1.07E-02
K00821		00300 Lysine biosynthesis	acetylornithine/N-succinyldiaminopimelate aminotransferase [EC:2.6.1.11]	20	11	2.30	1.34E-05	2.40E-03
K01636		00350 Tyrosine metabolism	hexulose-6-phosphate synthase [EC:4.1.2.-] [EC:4.1.2.43]	2	46	-3.08	1.09E-04	1.07E-02
K03119	01160 Metabolism of Other Amino Acids	00430 Taurine and hypotaurine metabolism	taurine dioxygenase [EC:1.14.11.17]	5	0	NA	1.42E-03	4.64E-02
K01925	01170 Glycan Biosynthesis and Metabolism	00550 Peptidoglycan biosynthesis	UDP-N-acetylmuramoylalanine--D-glutamate ligase [EC:6.3.2.9]	9	3	3.02	7.18E-04	3.68E-02
K00120	01195 Biosynthesis of Secondary Metabolites	00903 Limonene and pinene degradation	putative glucose dehydrogenase precursor [EC:1.1.-.-]	9	2	3.61	2.36E-04	1.89E-02
K00257	01196 Xenobiotics Biodegradation and Metabolism	00281 Geraniol degradation	acyl-CoA dehydrogenase	56	70	1.12	2.93E-05	3.71E-03
K00001		00980 Metabolism of xenobiotics by cytochrome P450	alcohol dehydrogenase [EC:1.1.1.1]	9	92	-1.91	8.91E-06	1.75E-03
K02601	01210 Transcription	03028 Other transcription related proteins	transcriptional antiterminator NusG	12	83	-1.35	1.08E-03	3.84E-02
K02888	01220 Translation	03010 Ribosome	large subunit ribosomal protein L21	5	73	-2.43	8.28E-06	1.75E-03
K02926			large subunit ribosomal protein L4	61	90	0.88	3.20E-04	2.30E-02
K02931			large subunit ribosomal protein L5	45	267	-1.13	1.72E-07	7.42E-05
K02935			large subunit ribosomal protein L7/L12	49	220	-0.73	9.19E-04	3.84E-02
K02939			large subunit ribosomal protein L9	12	94	-1.53	1.10E-04	1.07E-02
K02945			small subunit ribosomal protein S1	47	223	-0.81	2.58E-04	1.98E-02
K02982			small subunit ribosomal protein S3	74	118	0.77	4.39E-04	2.70E-02
K03386	01230 Folding, Sorting and Degradation	03100 Protein folding and associated processing	peroxiredoxin (alkyl hydroperoxide reductase subunit C) [EC:1.11.1.15]	7	62	-1.71	9.63E-04	3.84E-02
K05838			putative thioredoxin	7	0	NA	1.03E-04	1.07E-02
K03582	01240 Replication and Repair	03034 Other replication, recombination and repair proteins	exodeoxyribonuclease V beta subunit [EC:3.1.11.5]	6	0	NA	3.83E-04	2.50E-02
K02036	01310 Membrane Transport	02010 ABC transporters	phosphate transport system ATP-binding protein [EC:3.6.3.27]	6	55	-1.76	1.38E-03	4.58E-02
K02037			phosphate transport system permease protein	11	111	-1.90	1.10E-06	3.94E-04
K02040			phosphate transport system substrate-binding protein	25	305	-2.17	7.73E-19	1.67E-15
K02055			putative spermidine/putrescine transport system substrate-binding protein	1	44	-4.02	2.57E-05	3.71E-03
K02407	01410 Cell Motility	02040 Flagellar assembly	flagellar hook-associated protein 2	8	76	-1.81	1.12E-04	1.07E-02
K03529	01420 Cell Growth and Death	04410 Cell division	chromosome segregation protein	17	8	2.53	2.17E-05	3.59E-03

*KO = KEGG ortholog number; Control and DOM are raw counts of sequences annotated as a KEGG ORF in the controls and treatments. Note for this analysis, all controls were pooled based on the results of the ANOVA in Supplemental Table; ln(Fold change) is the natural log of the estimated fold change of the DOM12 relative to the DOM27 treatment (i.e., positive values indicate enrichment in the DOM12 treatment). The fold changes are calculated after scaling by the number of non-rRNA reads in the library (see Methods); *q*-value is a calibration of the table-wide false discovery rate (Storey et al., 2003). Note that some KEGG othologs belong to multiple functional categories and pathways. For brevity, we have included only one designation for each ortholog.

Supplemental Table S7: Poisson ANOVA of RuMP Pathway*

EC Number	AIC	Intercept Coefficient	DOM Coefficient	P-value	q-value	T2 Coefficient	P-value	q-value	T12 Coefficient	P-value	q-value	T27 Coefficient	P-value	q-value
1.1.1.44	25.32	-9.82	**1.42**	**3.96E-03**	**7.92E-03**	**-20.80**	**1.27E-03**	**3.27E-03**	**1.12**	**2.74E-03**	**1.23E-02**	-0.16	8.94E-01	1.00E+00
3.1.1.31	19.94	-30.12	1.48	2.39E-02	3.92E-02	-0.60	4.39E-02	7.90E-02	19.90	6.69E-01	8.01E-01	19.68	3.58E-01	8.30E-01
1.1.1.49	25.80	-9.82	-0.60	5.55E-01	6.24E-01	**-20.71**	**3.32E-03**	**7.48E-03**	1.78	3.27E-01	5.35E-01	1.58	8.09E-02	3.64E-01
5.3.1.9	38.19	-9.82	**1.15**	**1.80E-03**	**4.51E-03**	-0.08	1.78E-01	2.67E-01	0.56	7.48E-01	8.01E-01	0.77	4.70E-01	9.40E-01
5.3.1.9	12.00	-35.12	-23.75	2.12E-01	2.54E-01	1.50	4.99E-01	6.42E-01	25.65	1.02E-01	2.29E-01	0.98	1.00E+00	1.00E+00
4.2.1.12	38.72	-9.82	**1.45**	**4.00E-07**	**1.80E-06**	**-0.72**	**9.65E-04**	**2.89E-03**	0.47	4.93E-02	1.48E-01	1.28	1.90E-01	6.84E-01
2.7.6.1	37.81	-7.62	**0.94**	**6.70E-05**	**2.41E-04**	**-1.42**	**1.82E-04**	**8.18E-04**	-0.08	8.62E-01	8.62E-01	-0.15	7.39E-01	1.00E+00
2.2.1.2	67.25	-7.87	0.50	3.37E-02	5.05E-02	**-1.55**	**3.17E-04**	**1.14E-03**	0.01	7.06E-01	8.01E-01	0.15	7.60E-01	1.00E+00
4.3.-.-	27.41	-28.12	**1.05**	**2.01E-03**	**4.51E-03**	17.21	4.08E-02	7.90E-02	17.85	6.10E-02	1.57E-01	19.16	6.35E-02	3.64E-01
2.3.1.101	13.94	-35.12	**24.10**	**1.73E-03**	**4.51E-03**	-22.58	5.77E-02	9.43E-02	-22.63	2.50E-02	9.00E-02	2.44	1.00E+00	1.00E+00
3.5.4.27	12.61	-35.12	23.28	1.17E-01	1.52E-01	-21.99	3.43E-01	4.74E-01	-22.04	2.62E-01	4.72E-01	1.87	1.00E+00	1.00E+00
5.1.3.1	30.12	-9.82	0.76	2.37E-02	3.92E-02	0.50	5.88E-01	7.06E-01	0.67	7.56E-01	8.01E-01	0.92	3.69E-01	8.30E-01
2.2.1.1	53.02	-6.73	0.26	1.18E-01	1.52E-01	-0.17	6.93E-01	7.79E-01	-0.32	1.50E-01	3.00E-01	-0.03	9.04E-01	1.00E+00
2.7.9.2	32.35	-31.12	**4.45**	**2.17E-75**	**3.91E-74**	**19.77**	**1.07E-13**	**1.92E-12**	**21.65**	**1.01E-04**	**6.06E-04**	21.16	2.30E-01	6.90E-01
1.5.1.5	34.48	-9.82	-0.17	6.60E-01	6.60E-01	1.63	7.52E-01	7.96E-01	1.25	5.09E-01	7.64E-01	1.81	3.23E-02	3.64E-01
5.-.-.-	18.37	-36.12	**27.02**	**6.53E-27**	**5.87E-26**	**-24.69**	**7.66E-11**	**6.89E-10**	**-0.57**	**1.58E-12**	**2.84E-11**	2.97	1.00E+00	1.00E+00
4.1.2.43	18.28	-35.12	**25.98**	**1.65E-14**	**9.89E-14**	**-23.67**	**3.32E-06**	**1.99E-05**	**0.16**	**2.86E-05**	**2.58E-04**	2.30	1.00E+00	1.00E+00
4.1.2.14	28.03	-28.12	-0.11	5.90E-01	6.25E-01	19.06	8.04E-01	8.04E-01	18.61	6.23E-01	8.01E-01	19.26	4.65E-02	3.64E-01

* EC Numbers correspond to orthologs in Figure 3; AIC = Akaike Information Criterion; *q*-value is a calibration of the table-wide false discovery rate (Storey et al., 2003); See methods for definitions of the coefficients.

Supporting Information References

1. Frias-Lopez J, *et al.* (2008) Microbial community gene expression in ocean surface waters. *Proc. Natl. Acad. Sci, U.S.A.* 105: 3805-3810.
2. DeLong EF, *et al.* (2006) Community Genomics Among Stratified Microbial Assemblages in the Ocean's Interior. *Science* 311:496-503.
3. Marie D, Partensky F, Jacquet S, & Vaulot D (1997) Enumeration and Cell Cycle Analysis of Natural Populations of Marine Picoplankton by Flow Cytometry Using the Nucleic Acid Stain SYBR Green I. *Appl. Environ. Microbiol.* 63:186-193.
4. Rodrigue Sb, *et al.* (2009) Whole Genome Amplification and *De novo* Assembly of Single Bacterial Cells. *PLoS ONE* 4(9):e6864.
5. DeSantis TZ, *et al.* (2006) Greengenes, a Chimera-Checked 16S rRNA Gene Database and Workbench Compatible with ARB. *Appl. Environ. Microbiol.* 72:5069-5072.
6. Wang Q, Garrity GM, Tiedje JM, & Cole JR (2007) Naive Bayesian Classifier for Rapid Assignment of rRNA Sequences into the New Bacterial Taxonomy. *Appl. Environ. Microbiol.* 73:5261-5267.
7. Pruesse E, *et al.* (2007) SILVA: a comprehensive online resource for quality checked and aligned ribosomal RNA sequence data compatible with ARB. *Nucl. Acids Res.* 35:7188-7196.
8. Ludwig W, *et al.* (2004) ARB: a software environment for sequence data. *Nucl. Acids Res.* 32:1363-1371.
9. Cole JR, *et al.* (2009) The Ribosomal Database Project: improved alignments and new tools for rRNA analysis. *Nucl. Acids Res.* 37(suppl_1):D141-145.
10. Letunic I & Bork P (2007) Interactive Tree Of Life (iTOL): an online tool for phylogenetic tree display and annotation. *Bioinformatics* 23:127-128.
11. Altschul SF, *et al.* (1997) Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. *Nucl. Acids Res.* 25:3389-3402.
12. Huson DH, Auch AF, Qi J, & Schuster SC (2007) MEGAN analysis of metagenomic data. *Genome Research* 17:377-386.
13. Urich T, *et al.* (2008) Simultaneous Assessment of Soil Microbial Community Structure and Function through Analysis of the Meta-Transcriptome. *PLoS ONE* 3:e2527.
14. Wang L, Feng Z, Wang X, Wang X, & Zhang X (2010) DEGseq: an R package for identifying differentially expressed genes from RNA-seq data. *Bioinformatics* 26:136-138.
15. Shi Y, Tyson GW, & DeLong EF (2009) Metatranscriptomics reveals unique microbial small RNAs in the ocean's water column. *Nature* 459:266-269.
16. Li X & Qin L (2005) Metagenomics-based drug discovery and marine microbial diversity. *Trends Biotechnol.* 23:539-543.
17. Wang L, Feng Z, Wang X, Wang X, & Zhang X (2010) Degseq: an R package for identifying differentially expressed genes from RNA-seq data. 26:136-138.
18. Kristiansson E, Hugenholtz P, & Dalevi D (2009) ShotgunFunctionalizeR: An R-package for functional comparisons of metagenomes. *Bioinformatics* 25:2737-2738.

19. Team RDC (2010) R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing. Version 2.11.1 (2010-05-31).
20. Storey JD (2003) The positive false discovery rate: A Bayesian interpretation and the q-value. *Ann Stat.* 31: 2013-2035.

Appendix B: Time-series analyses of Monterey Bay coastal microbial picoplankton using a 'genome proxy' microarray

Virginia I. Rich, Vinh D. Pham, John Eppley, Yanmei Shi, and Edward F. DeLong

Rich, V. I., Pham, V. D., Eppley, J., Shi, Y. and DeLong, E. F., Time-series analyses of Monterey Bay coastal microbial picoplankton using a 'genome proxy' microarray. *Environmental Microbiology*, doi: 10.1111/j.1462-2920.2010.02314.x

environmental microbiology

Environmental Microbiology (2010)

doi:10.1111/j.1462-2920.2010.02314.x

Time-series analyses of Monterey Bay coastal microbial picoplankton using a 'genome proxy' microarray

Virginia I. Rich,† Vinh D. Pham, John Eppley, Yanmei Shi and Edward F. DeLong*

Department of Civil and Environmental Engineering, Massachusetts Institute of Technology, 48-427, 15 Vassar Street, Cambridge, MA 02139, USA.

Summary

To investigate the temporal, spatial and phylogenetic resolution of marine microbial community structure and variability, we designed and expanded a genome proxy array (an oligonucleotide microarray targeting marine microbial genome fragments and genomes), evaluated it against metagenomic sequencing, and applied it to time-series samples from the Monterey Bay. The expanded array targeted 268 microbial genotypes across much of the known diversity of cultured and uncultured marine microbes. The target abundances measured by the array were highly correlated to pyrosequence-based abundances (linear regression $R^2 = 0.85–0.91$, $P < 0.0001$). Fifty-seven samples from ~4 years in Monterey Bay were examined with the array, spanning the photic zone (0 m), the base of the surface mixed layer (30 m) and the subphotic zone (200 m). A significant portion of the expanded genome proxy array's targets showed signal (95 out of 268 targets present in ≥ 1 sample). The multi-year community survey showed the consistent presence of a core group of common and abundant targeted taxa at each depth in Monterey Bay, higher variability among shallow than deep samples, and episodic occurrences of more transient marine genotypes. The abundance of the most dominant genotypes peaked after strong episodic upwelling events. The genome-proxy array's ability to track populations of closely related genotypes indicated population shifts within several abundant target taxa, with specific populations in some cases clustering by depth or oceanographic season. Although 51 cultivated organisms were targeted (representing 19% of the array) the majority of targets detected and of total target signal (85% and ~92% respectively) were from uncultivated genotypes, often those derived from Monterey Bay. The array provided a relatively cost-effective approach (~$15 per array) for surveying the natural history of uncultivated lineages.

Introduction

Marine microbial communities are major drivers in global biogeochemical cycling (Arrigo, 2005; Howard *et al.*, 2006; Karl, 2007), sources of metabolic discoveries (e.g. (Béjà *et al.*, 2000; Kolber *et al.*, 2000; Dalsgaard *et al.*, 2003; Kuypers *et al.*, 2003), and the focus of metagenomic surveys beyond the scale of those yet undertaken in other habitats (Venter *et al.*, 2004; Tringe *et al.*, 2005; DeLong *et al.*, 2006; Kennedy *et al.*, 2007; Rusch *et al.*, 2007; Wegley *et al.*, 2007; Wilhelm *et al.*, 2007; Yooseph *et al.*, 2007; Dinsdale *et al.*, 2008; Marhaver *et al.*, 2008; Mou *et al.*, 2008; Neufeld *et al.*, 2008). However, microbial community dynamics remain poorly understood due to technical limitations and the analytical challenges of high-resolution spatial and temporal studies. Most studies capture spatiotemporal snapshots or focus on one or a few groups over space and time. While the value of improved resolution is clear, lower-resolution (e.g. in time, space or diversity of target organisms) studies have provided much insight into microbial community variability over space and time. For example, such studies reveal changing community structure that correlates to environmental parameters, and even climate change responses [e.g. Hawaii Ocean Time Series (Karl, 1999; Karner *et al.*, 2001), Bermuda Atlantic Time Series (Morris *et al.*, 2005) and San Pedro Ocean Time-Series (Fuhrman *et al.*, 2006)].

To gain a higher-resolution picture of microbial community variability, we developed the 'genome proxy' array (Rich *et al.*, 2008) which uses sets of multiple, distributed 70-mer probes to target genotypes (genome fragments and genomes) as a cost-effective high-throughput survey tool to track microbial community variability. The array cross-hybridizes to related genotypes that approach

Received 20 January, 2010; accepted 20 June, 2010. *For correspondence. E-mail delong@mit.edu; Tel. (+1) 617 253 5271; Fax (+1) 617 253 2679. †Present address: Department of Ecology and Evolutionary Biology, University of Arizona, 1041 East Lowell Street, Tucson, AZ 85721, USA.

≥ ~80% average nucleotide identity (ANI, as in Konstantinidis and Tiedje, 2005), with the stringency and specificity adjustable *in silico* to ≥ ~90% ANI. Related cross-hybridizing strains produced distinct hybridization patterns across their target probe set, and the array can thereby reveal shifts in population structure across samples (Rich *et al.*, 2008). The limit of detection is approximately 0.1% of the community for targeted genotypes, and approximately 1% of the community for related, cross-hybridizing genotypes (Rich *et al.*, 2008).

We report here on an expanded genome proxy array that targets 268 genotypes (from 14 in the original). We ground-truthed the array signal using pyrosequenced community DNA, and applied the optimized array to investigate the time-series microbial dynamics over a 4-year period at Monterey Bay Station M1 (36.747°N, 122.022°W). This microbially and oceanographically well-studied coastal environment (e.g. Pennington and Chavez, 2000; Suzuki *et al.*, 2001a,b; 2004; O'Mullan and Ward, 2005; Ward, 2005; Mincer *et al.*, 2007; Pennington *et al.*, 2007) is characterized by strong seasonal upwelling, providing a contextually rich first real-world application of this tool. In all, we hybridized 57 archived DNA samples collected over 4 years from oceanographic water column features (photic, base of the mixed layer and subphotic) to identify patterns in and drivers of microbial community structure.

Results and discussion

Development and ground-truthing of the expanded genome proxy array

The expanded genome proxy array targets 268 microbial genotypes, through suites of probes (~20 per target) dispersed along genomes and genome fragments derived from microbes inhabiting marine habitats. Targeted organisms were selected to span known marine microbial diversity (16S rRNA-containing targets are shown in Fig. 1 and Figs S1–S5, all targets are listed in Table S1 and summarized in Table S2). For diverse and abundant marine clades, representatives were chosen where possible from each known lineage and from multiple geographic origins.

We compared the results from the expanded array to those obtained using pyrosequencing of the same microbial community DNA for three different Monterey Bay surface samples [Julian Day (JD) 298 in 2000, and JD115 and JD135 in 2001]. A full GS-FLX pyrosequencing run (~400 000 reads) was performed per sample, trimmed to remove poor quality sequence (~5.5% of reads), and 'hybridized' *in silico* using BLAST (Altschul *et al.*, 1990) to the 268 genotypes targeted by the array. To simulate the amount of sequence divergence tolerated by the array, BLAST parameters were calibrated using array results for genomes of related *Prochlorococcus* strains whose relative cross-hybridization to the array had been experimentally determined (Rich *et al.*, 2008). Using this approach (see *Experimental procedures*), 1.9–2.5% of the total pyrosequencing reads in these three samples were assigned to array targets (7636/395767 for 0m_2000-298, 8743/345650 for 0m_2001-115 and 9252/39197 for 0m_2001-135), of which ~66–75% were assigned to only 12 targets in all three samples. Eleven of these 12 targets were environmental genomic clones (predominantly from the SAR86 and *Roseobacter* clades) while the tenth was the genome of a cultured NAC11-7 clade *Roseobacter*.

The normalized pyrosequencing read recruitment was strongly correlated to the normalized unfiltered mean array intensity (linear regression with R^2 of 0.85–0.91 across three samples, P-values < 0.0001; Fig. 2). Such strong correlation between the relatively unbiased (no cloning biases, etc.) direct pyrosequencing method and the high-throughput genome proxy array provided support for the veracity of the array as a tool for profiling studies requiring high sample throughput.

Exploring microbial communities using the genome proxy array

We hybridized community DNA from 57 Monterey Bay samples at Station M1 over 4 years (sample overview in Fig. 3) to the expanded genome proxy microarray. Approximately one-third of the array's diverse targets (95 of 268 targets) were present in one or more of the samples at this site. To be considered present, a target was required to show signal in > 40% of its probes, to avoid single-probe high-identity cross-hybridizations from unrelated taxa (as empirically determined in Rich *et al.*, 2008, see *Experimental procedures*). The majority of targets detected by array were uncultivated marine lineages, many of which originated from Monterey Bay (Fig. S6A).

Shallow versus deep profiles. Hierarchical clustering (Fig. 4) and canonical discriminant analyses (CDA, Fig. 5) revealed clear community structure throughout the oceanographic depth profiles sampled, with greater variability among shallow samples than deep ones (see branch lengths of hierarchical clustering and intensity of array signals in Fig. 4). For example, the Monterey Bay surface photic-zone samples (0 and 30 m) were less similar to each other (as indicated by branch distances) than the subphotic-zone samples were to one another (200 m, Figs 4 and 5). Depth-structuring in microbial populations and communities is well described in

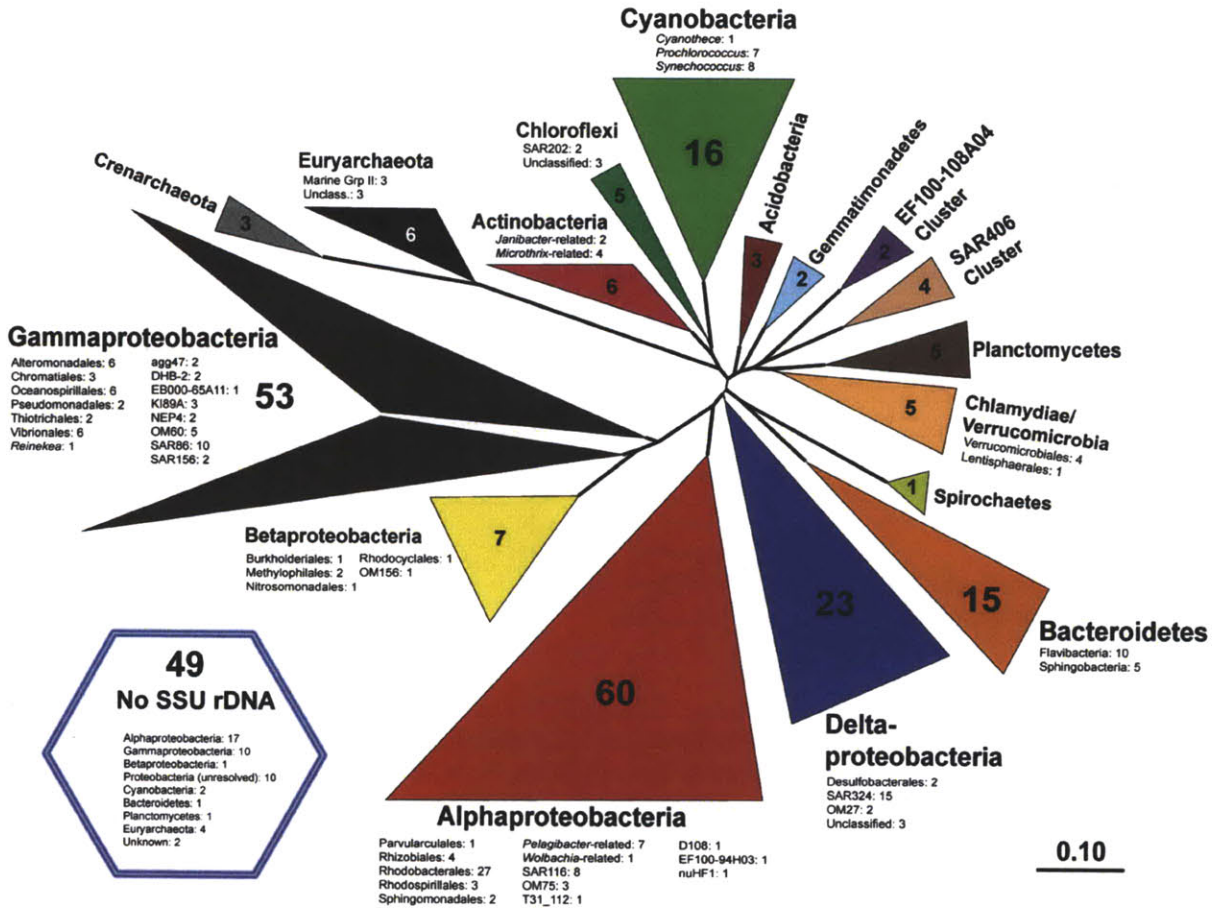


Fig. 1. Radial tree illustrating the phylogenetic relationships among the 268 targets of the expanded genome proxy array. Numbers indicate the number of targets within each phylogenetic clade. Sequences from clones lacking a small-subunit rRNA gene (SSU) phylomarker are represented separately by the hexagon. Tree was created based on alignment of 16S rRNA sequences using the SILVA database Release 99 (Pruesse *et al.*, 2007) with the ARB software package (Ludwig *et al.*, 2004).

marine systems at the level of rRNA profiling (e.g. Fuhrman *et al.*, 1992; Field *et al.*, 1997; Karner *et al.*, 2001; Bano and Hollibaugh, 2002; Morris *et al.*, 2004; Suzuki *et al.*, 2004; Treusch *et al.*, 2009) and fosmid end-sequencing (DeLong *et al.*, 2006), so it is not surprising that our genome proxy array reveals similar structure with respect to the targeted community genotypes examined here. These differential depth distributions extended to the majority of observed taxa, with four notable depth-specific groups of targets (dashed boxes in Fig. 4 and detailed in Table 1). Eight targets were present in >90% of shallow samples ('shallow-consistent'), 10 were present in 50–90% of shallow samples ('shallow-frequent'), 10 were present in >90% of deep samples ('deep-consistent'), and three were present in 50–90% of deep samples ('deep-frequent') (Table 1). Notably, the differential presence and distribution of three to five targeted genotypes in each depth drove the three depth's separation of array profiles (CDA, Fig. 5A).

While there was clear photic versus subphotic depth structure, the 0 m and 30 m array profiles were intermingled despite their generally different chemical and physical environments (Fig. 3). While we selected 30 m as the base of the mixed layer to attempt to capture the nitricline, it is clear that the mixed layer depth (MLD) at this site usually lacks a discrete thermocline and moves dramatically over short time periods (see calculated MLD across sampling dates, Fig. S7). Therefore, our sampling strategy might have been improved by varying sampling depths based on calculated single-time-point MLDs for each cruise; however, removing 30 m samples that were clearly above the MLD and reclustering the array profiles did not resolve samples into 0 m and 30 m clusters (Fig. S8), emphasizing the highly dynamic nature of these photic-zone waters.

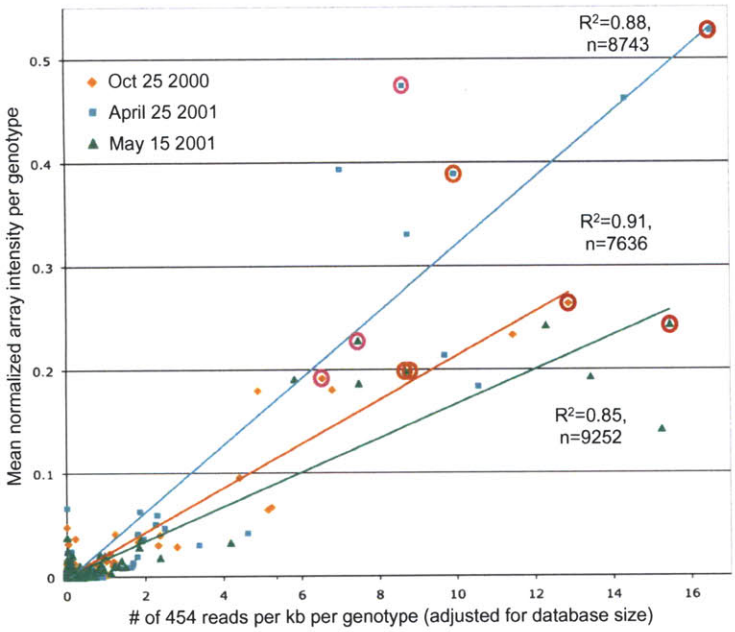


Fig. 2. Cross-comparison of array- and pyrosequence-based target abundances for three MB samples. The *P*-values associated with each linear regression were < 0.0001, and the R^2 values and number of recruited pyrosequences are indicated. Using BLASTN parameters optimized to mimic array cross-hybridization, all 268 targeted genomes and genome fragments were compared (using BLAST) to the pyrosequence data derived from identical samples. Pyrosequences were assigned to one or more array targets, proportional to the bitscore of each match. The number of pyrosequences matching each target was normalized to target length and database size, and compared with the unfiltered array signal (see *Experimental procedures* and *Results*) of the same clone. Correlation lines were not forced through the origin. Circled data points indicate proteorhodopsin-containing clones abundant by array signal post-upwelling as described in the text: red circles = EB000-55B11, orange circles = EB000-39F01, and pink circles = *Rhodobacterales* HTCC2255.

Profile correlations to ocean chemistry. Array-based sample profiles compared between depths were strongly correlated to each tested nutrient as follows: phosphate, nitrate and silicate drove the differentiation of the shallow from the deep samples, while nitrite drove the separation of 30 m from 0 m (Fig. 5B). Samples from each depth were separately subjected to PCA (Fig. 6), indicating that nutrients did not separate the 0 m samples (Fig. 6A), but were important at both 30 m and 200 m. Specifically, at 30 m (Fig. 6B), nutrient variability was correlated to the principal component axes, with a strong upwelling signal of phosphate, nitrate and silicate and a slightly weaker and inverse signal for nitrite (likely from remineralization). Finally, at 200 m (Fig. 6C), nitrate and nitrite showed no and weak correlations, respectively, while silicate and phosphate gave strong but non-overlapping correlations. Overall, these correlations to nutrient concentrations recapitulate the oceanographic differences in nutrients with depth at this location (Fig. 3).

Tracking abundant taxa. Not surprisingly, one of the most commonly detected bacterial groups was the *Roseobacter* clade (Fig. 4). This metabolically diverse group commonly comprises up to 20% of cells in coastal waters (reviewed in Buchan *et al.*, 2005), including high abundances (20–40% of rRNA clone libraries) in the mid-Monterey Bay region during upwelling (Suzuki *et al.*, 2001b). More specifically, in fosmid clone libraries from Monterey Bay the *Roseobacter* NAC11-7 and CHAB-I-5 clades comprised nearly 30% of the 16S-containing clones (27% and 29% at 0 and 80 m respectively) and ~80% of the total *Roseobacter* signal at 0 and 80 m, while at 100 m NAC11-7 disappeared and CHAB-1-5 persisted at low abundance (Suzuki *et al.*, 2004) (see Table S3 for clade-by-clade comparison of array results with previous Monterey Bay community surveys). In agreement with these previous single-time-point observations, the array profiles indicate high *Roseobacter* abundances over time (Fig. 4 and Fig. S9A). Twenty-eight per cent of the commonly occurring targeted taxa in surface waters were NAC11-7 clones (four of eight targets in the *shallow-consistent* group, and 1 of 10 *shallow-frequent* group; listed in Table 1), and 1 of the 10 *deep-consistent* taxa was a CHAB-I-5 clone (Table 1). In addition, another CHAB-I-5 clone (EB080_L58F04) was present in 35% of shallow samples. Further, differential NAC11-7 distributions drove the differentiation of 30 m from 0 m samples (three of five driving taxa, Fig. 5A).

A second abundant shallow water bacterial group was the uncultivated gammaproteobacterial SAR86 clade, which is commonly reported in marine samples (Eilers *et al.*, 2000; Rappé *et al.*, 2000; Suzuki *et al.*, 2001b; Venter *et al.*, 2004; Morris *et al.*, 2006), known to partition with depth (Morris *et al.*, 2006), and can comprise up to 10% of the cells in a community (Mullins *et al.*, 1995; Eilers *et al.*, 2000; Morris *et al.*, 2006). In Monterey Bay, it is abundant in rRNA clone libraries during upwelling (3–6% of total bacterial SSU DNAs; Suzuki *et al.*, 2001b), and in large-insert clone libraries (5.6%, 5.5% and 1.6%, respectively, of the SSU operon-containing clones 0 m, 80 m and 100 m; Suzuki *et al.*, 2004; Table S3). Array-based profiling reflected also this high SAR86 abundance

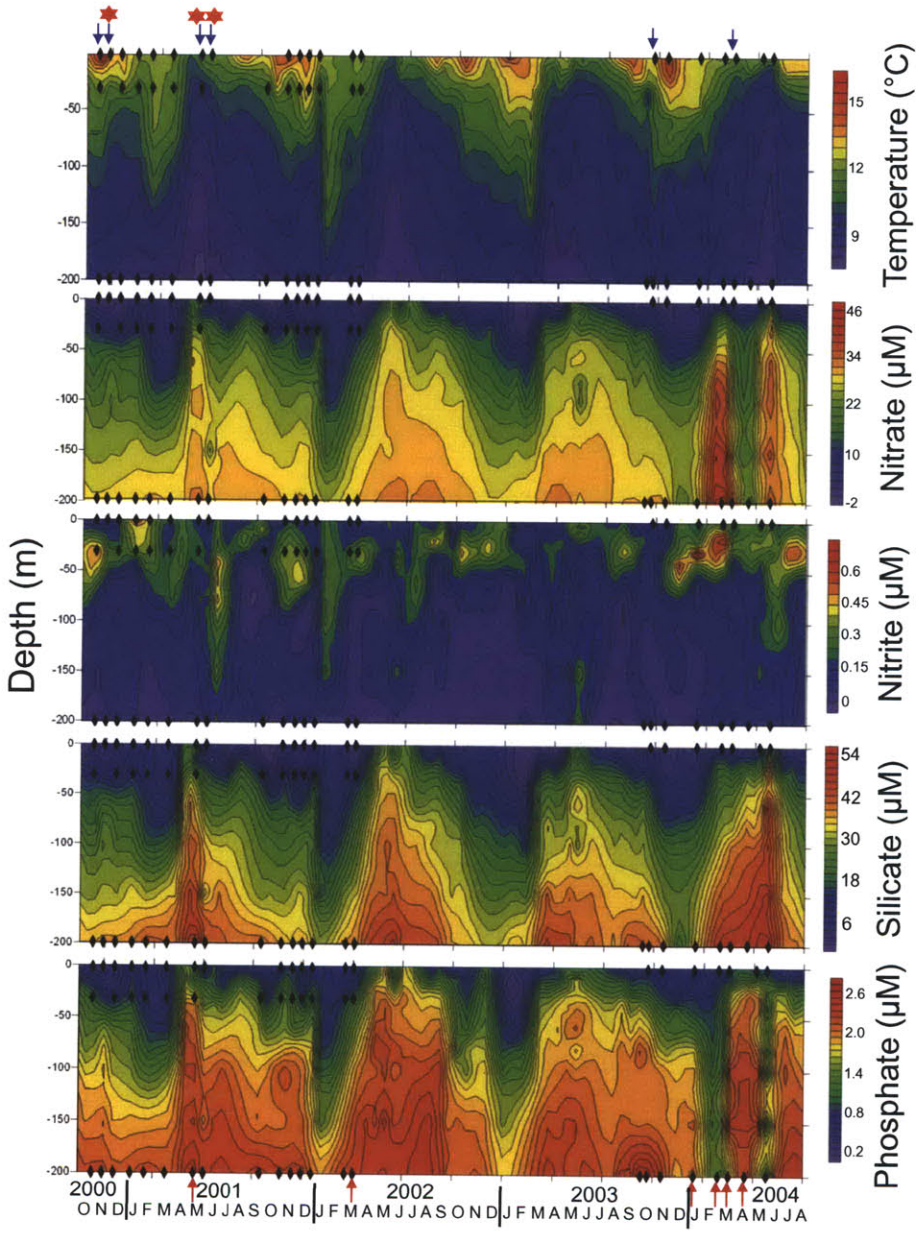


Fig. 3. Sample origin from Monterey Bay Station M1 over depth (*y*-axis) and time (*x*-axis) against the backdrop of oceanographic context. The 57 samples (black diamonds) hybridized to the array derive from three depths (0, 30 and 200 m) over ~4 years; time (with months indicated by their first-letter designations) is indicated along the *x*-axis. The 0 m samples used for cross-validation pyrosequencing are indicated by red stars. Panels show temperature, nitrate, nitrite, silicate and phosphate concentrations. Blue arrows at top of each panel indicate samples whose 0 m array profiles were particularly intense. Red arrows at bottom of panels indicate 200 m samples whose variability was correlated to silicate and phosphate.

(Fig. 4 and Fig. S9B); 22% of common shallow water targets (two *shallow-consistent* and two *shallow-frequent*) were SAR86 clones. The distribution of one particular SAR86 target (a Monterey-derived environmental clone) helped drive the differentiation of 30 m samples from those at 0 m (Fig. 5A).

A remaining *shallow-frequent* target of note was an alphaproteobacterial SAR116-I clone. Of 12 SAR116 targets, two originated in Monterey Bay, and these were the only phylotypes detected (Fig. 4). The SAR116-II target was present only twice, in 0 m samples, while the SAR116-I clone was present in 62% of shallow samples.

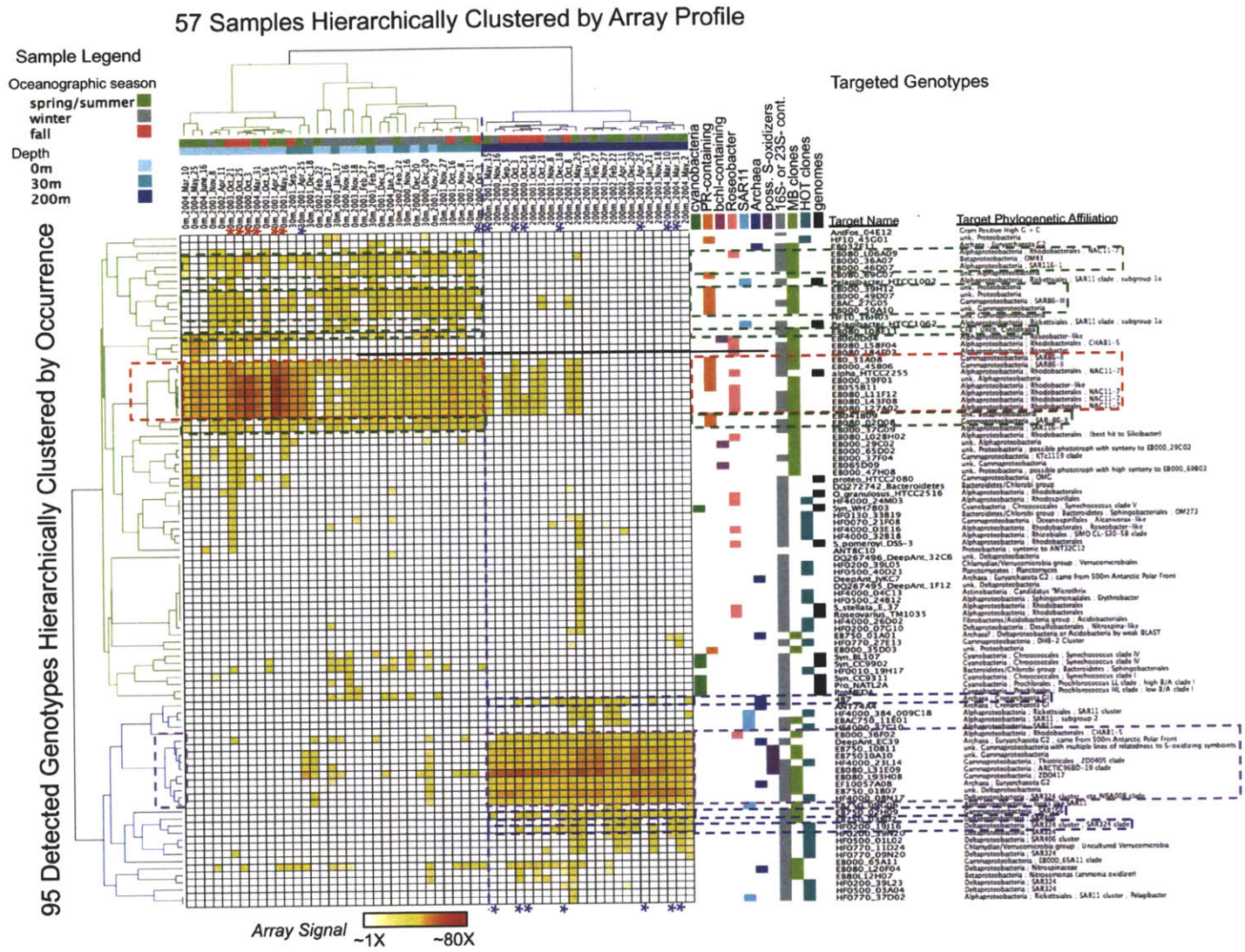


Fig. 4. Clustering of hybridizations by sample and by genotype. Hierarchical clustering was performed in GenePattern using Pearson correlation (see *Experimental procedures*) and is sh across the top for samples and along the side for genotypes. Targets are colour-coded by phylogenetic identity, gene content of particular interest (note column indicating presence/absen of 16S rRNA gene), and origin (see colour legend; MB = Monterey Bay, HOT = Hawaii Ocean Time series). Intensity of yellow-to-red colour for each genotype and sample date indicates relative target signal; note that relative abundance is quantitative for each genotype between samples but not between genotypes. Samples are named Depth_Year_CollectionDate, and colour-coded by depth and by oceanographic season (see colour legend and text). The break between shallow and deep clusters is indicated by the blue vertical dashed line. Abundant targets referred to in the text are boxed with dashed lines, 'shallow-consistent' = red, 'shallow-frequent' = green, 'deep-consistent' = purple, 'deep-frequent' = navy. Red asterisks denote samples with particularly intense 0 m profiles; the 30 m and 200 m samples for the same dates, when available, are indicated by blue asterisks.

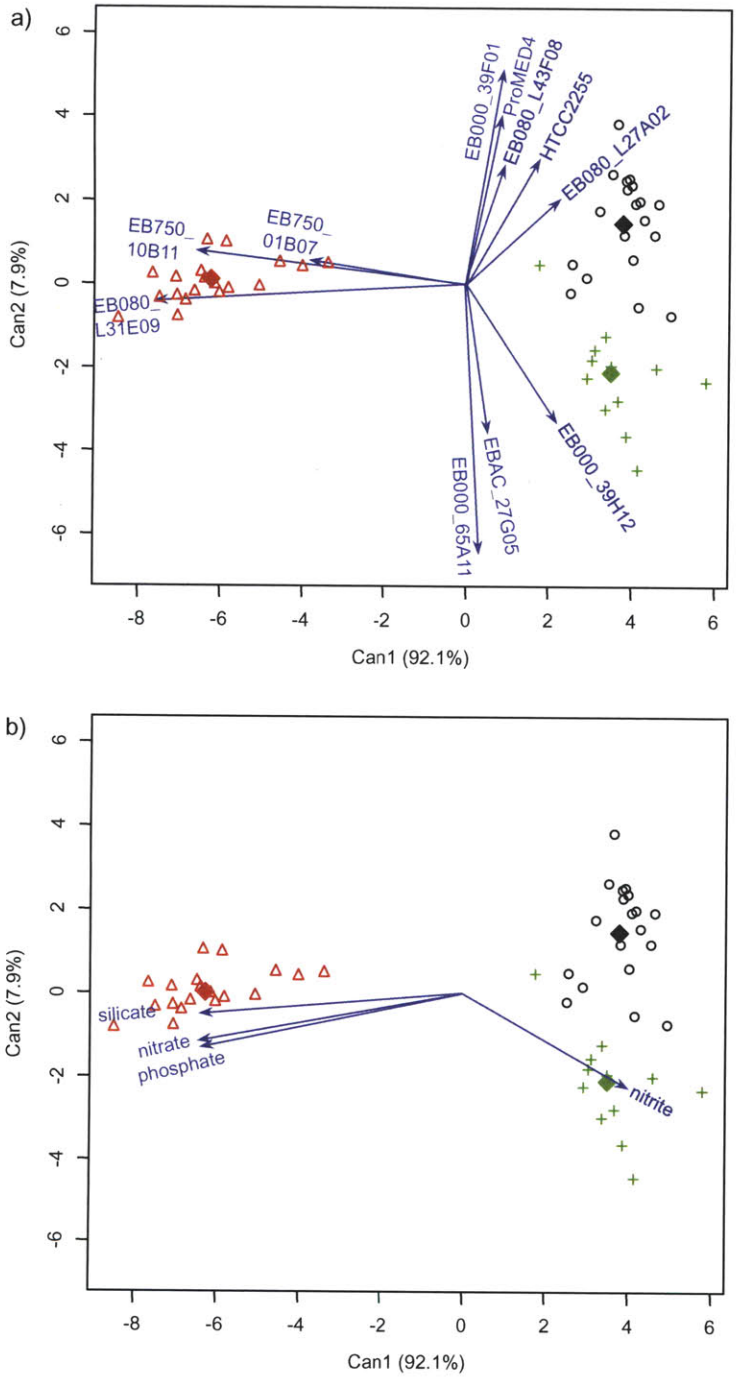


Fig. 5. Canonical discriminant analysis (c.d.) of Monterey Bay sample (0 m O, 30 m +, and 200 m Δ) array data, with parameter correlations to c.d. axes indicated by vector length and direction. Diamonds designate centre of each depth's data cloud.
A. Genotype abundance correlations to c.d. axes; the distribution of particular taxa drive the differentiation of depths.
B. Nutrient correlations to c.d. axes; nutrients are dramatically different between the three depths, and this strong difference is recapitulated in the correlations to c.d. axes.
Target taxonomic affiliations (by 16S identity, or by clone BLAST hits for clones with no 16S rRNA gene): EB000_39F01 = putative *Alphaproteobacteria*, ProMED4 = *Cyanobacteria;Prochlorococcus*, EB080_ L43F08 = *Alphaproteobacteria; Rhodobacterales;NAC11-7*, HTCC2255 = *Alphaproteobacteria;Rhodobacterales;NAC11-7*, EB080_L27A02 = *Alphaproteobacteria; Rhodobacterales;NAC11-7*, EB750_01B07 = putative *Deltaproteobacteria*, EB750_10B11 = *Gammaproteobacteria;related to S-oxidizing symbionts*, EB080_ L31E09 = *Gammaproteobacteria;ARCTIC96BD-19 clade, S-oxidizing symbiont relative*, EB000_ 39H12 = putative *Proteobacteria*, EBAC_ 27G05 = *Gammaproteobacteria;SAR86-III*, EB000_65A11 = *Gammaproteobacteria; EB000-65A11 clade.*

In large-insert environmental libraries from this site, the *Rhodospirillales* clade SAR116 comprised 11.3%, 1.4% and 0.8% of the SSU operon-containing clones in 0 m, 80 m and 100 m libraries respectively (Suzuki *et al.*, 2004; Table S3). The SAR116 clade has broad global distribution and frequently high abundances (e.g. Giovannoni and Rappé, 2000; DeLong *et al.*, 2006; Rusch *et al.*, 2007), but has only recently been isolated in culture (Stingl *et al.*, 2007). Due to the phylogenetic diversity of this clade (at least 10% divergent 16S rRNA, Stingl *et al.*, 2007), it is likely that the relative specificity of the array platform prohibited it from tracking other native but divergent SAR116 strains. The comparative array-versus-fosmid-libraries results suggest the need

Table 1. Array targets common in shallow or deep samples.

Category	Clone name[a]	Taxonomic identity	% Occurrence in shallow (0 m + 30 m)	% Occurrence in deep (200 m)
Shallow-consistent (present in 90–100% of samples, > 30 of out 34 samples)				
$n = 8$	EB000_31A08	*Proteobacteria*; *Gammaproteobacteria*; SAR86-II	100%	17%
	EB000_45B06	*Proteobacteria*; *Gammaproteobacteria*; SAR86-II	100%	22%
	EB000_55B11	*Proteobacteria*; *Alphaproteobacteria*; *Rhodobacter*-like	97%	30%
	EB080_L43F08	*Proteobacteria*; *Alphaproteobacteria*; *Rhodobacterales*; *Roseobacter* clade; NAC11-7	97%	35%
	EB080_L27A02	*Proteobacteria*; *Alphaproteobacteria*; *Rhodobacterales*; *Roseobacter* clade; NAC11-7	97%	35%
	alpha_HTCC2255	*Proteobacteria*; *Alphaproteobacteria*; *Rhodobacterales*; *Roseobacter* clade; NAC11-7	94%	30%
	EB080_L11F12	*Proteobacteria*; *Alphaproteobacteria*; *Rhodobacterales*; *Roseobacter* clade; NAC11-7	94%	35%
	EB000_39F01	Putative *Proteobacteria*; *Alphaproteobacteria*; (no 16S rRNA gene)	91%	30%
Shallow-frequent (present in 50–90% of samples, 17–30 out of 34 samples)				
$n = 10$	EB080_02D08	*Proteobacteria*; *Gammaproteobacteria*; SAR-86-II	85%	0%
	EB000_41B09	*Proteobacteria*; *Betaproteobacteria*[b]	82%	0%
	EB080_L06A09	*Proteobacteria*; *Alphaproteobacteria*; *Rhodobacterales*; *Roseobacter* clade; NAC11-7	79%	4%
	EB000_39H12	Putative *Proteobacteria*; (no 16S rRNA gene)	76%	0%
	EBAC_27G05	*Proteobacteria*; *Gammaproteobacteria*; SAR86-III	74%	9%
	EB000_36A07	*Proteobacteria*; *Betaproteobacteria*; OM43	68%	0%
	EB000_49D07	Putative *Proteobacteria*; (no 16S rRNA gene)	68%	9%
	EB080_L08E11	CFB; uncultivated Cytophaga	65%	0%
	EB000_46D07	*Proteobacteria*; *Alphaproteobacteria*; SAR116-1	62%	0%
	EB000_50A10	Putative *Proteobacteria*; *Gammaproteobacteria*; (no 16S rRNA gene)	59%	0%
Deep-consistent (present in 90–100% of samples, > 20 of 23 samples)				
$n = 10$	EB080_L31E09	*Proteobacteria*; *Gammaproteobacteria*; ARCTIC96BD-19 clade, S-oxidizing symbiont relative	29%	100%
	HF4000_23L14	*Proteobacteria*; *Gammaproteobacteria*; *Thiotricales*; ZD0405 clade	12%	100%
	EB750_10B11	Putative *Proteobacteria*; *Gammaproteobacteria*; (no 16S rRNA gene); carries RuBisCO gene and related to S-oxidizing symbionts[c]	9%	100%
	EB750_10A10	Putative *Proteobacteria*; *Gammaproteobacteria*; (no 16S rRNA gene); carries RuBisCO gene and related to S-oxidizing symbionts	9%	100%
	EB080_L93H08	*Proteobacteria*; *Gammaproteobacteria*; ZDO417	6%	100%
	EB750_01B07	*Proteobacteria*; *Deltaproteobacteria*	6%	100%
	HF4000_08N17	*Proteobacteria*; *Deltaproteobacteria*; SAR324 cluster; ctg_NISA008 clade	6%	100%
	EF100_57A08	*Archaea*; *Euryarchaeota*; Eury GII	3%	100%
	DeepAnt_EC39	*Archaea*; *Euryarchaeota*; Eury GII, came from 500 m Antarctic Polar Front	0%	100%
	EB000_36F02	*Proteobacteria*; *Alphaproteobacteria*; *Rhodobacterales*; *Roseobacter* clade; CHAB1-5	21%	96%
Deep-frequent (present in 50–90% of samples, 12–20 out of 23 samples)				
$n = 3$	EB750_02H09	*Proteobacteria*; *Gammaproteobacteria*; SAR156	0%	87%
	HF0200_19J16	*Proteobacteria*; *Deltaproteobacteria*; SAR324 cluster; SAR324 clade	0%	61%
	ORE_4B7	*Archaea*; *Crenarchaeota*; Cren GI	0%	57%

a. Clones with names beginning 'EB' or 'EF' originated from Monterey Bay, 'HF' from Hawaii, and the numbers preceding the underscore indicate depth of clone origin. See Table S1 for accession numbers and additional information.
b. Affiliation by phylogeny of three ribosomal proteins (McCarren and DeLong, 2007).
c. Affiliation by 30S ribosomal protein BLAST hit to *Vesicomyosocius* complete genome, and hits to S-oxidizing symbiont genes.

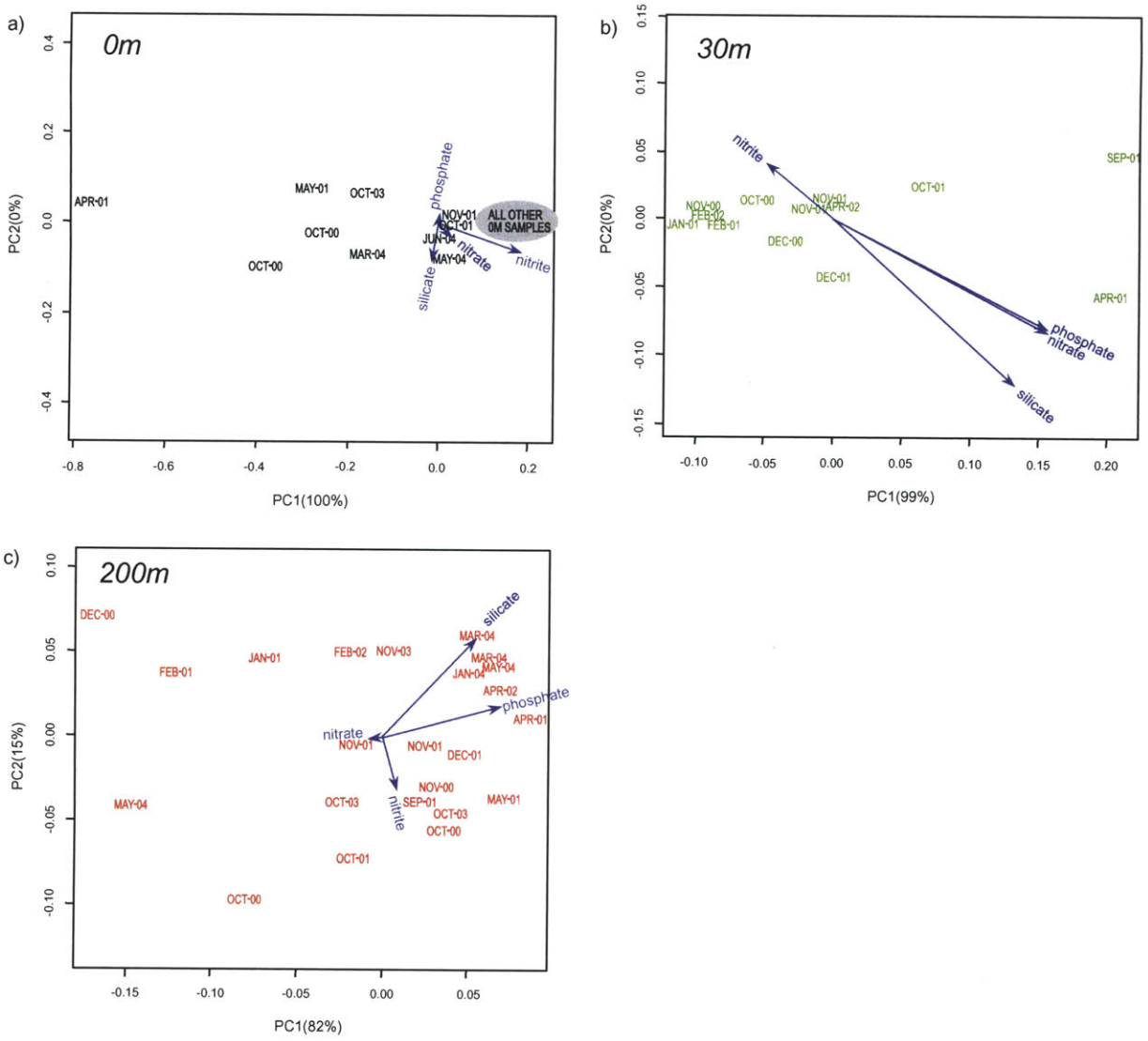


Fig. 6. Principal component (P.C.) analyses of Monterey Bay samples at each depth, with nutrient (nitrate, nitrite, phosphate and silicate) correlations to P.C. axes indicated by vector length and direction. Each sample is designated by its month and year.
A. Samples of 0 m; the sample variability among 0 m samples is not strongly correlated to differing nutrient concentrations.
B. Samples of 30 m; there is a strong correlation to all four nutrients, reflecting the upwelling signature at the base of the mixed layer.
C. Samples of 200 m; nitrite, phosphate and silicate each correlate to sample variability, in distinct ways.

for additional sequencing of environmental SAR116 genotypes.

Another common marine bacterial clade detected by the array was the alphaproteobacterial SAR11 clade, which is one of the most abundant heterotrophs in the global oceans (Morris *et al.*, 2002). Seven of the 10 targeted SAR11 genotypes were present in ≥ 1 Monterey Bay sample, and each showed depth-specific distribution (Fig. 4 and Fig. S9C). *Pelagibacter* HTCC1062 and HTCC1002, cultivated strains within the SAR11 subgroup 1a, were present only in shallow samples and occurred in ~30% of samples (29% and 35% respectively). Several other SAR11 environmental clone genotypes were present only in deep samples, and occurred frequently or sporadically. This is consistent with the known depth distributions of the two major SAR11 clades (Field *et al.*, 1997). Furthermore, the distribution of HTCC1062 and HTCC1002 showed no correlation to upwelling season, consistent with previous observations that their numbers do not change under phytoplankton bloom conditions (Morris *et al.*, 2005). The lower frequency of SAR11 genotypes than other clades,

combined with the clade's consistently high abundance measures by other methods, suggests the presence of many other SAR11 genotypes in these samples.

Targeted cyanobacteria did not show strong or consistent array signal in Monterey Bay. *Synechococcus* would be expected to be abundant in such nutrient-rich coastal waters (Waterbury *et al.*, 1986; Partensky *et al.*, 1999), and the array targeted eight marine *Synechococcus* across the group's known genomic diversity. The absence of strong cyanobacterial signal is therefore may be explained by the use of a 1.6 μm pre-filter during sample collection, which may have excluded larger *Synechococcus* cells (average uncultured cell size 0.8–2.2 μm, Waterbury *et al.*, 1979). Both *Synechococcus* and *Prochlorococcus* were sporadically detected in surface waters (Fig. 4), and the differential distribution of *Prochlorococcus* MED4 helped differentiate 0 m from 30 m samples (Fig. 5A).

The array captured information about *deep-consistent* genotypes (Fig. 4, Table 1) including four gammaproteobacterial targets (EB080_L31E09, EB750-10B11, EB750-10A10 and HF4000-23L14) related to chemoautotrophic deep-sea invertebrate symbionts and commonly observed in water column 16S rRNA surveys (López-García *et al.*, 2001; Bano and Hollibaugh, 2002; Zubkov *et al.*, 2002; Klepac-Ceraj, 2004; Suzuki *et al.*, 2004; Stevens and Ulloa, 2008; Walsh *et al.*, 2009), one of which (EB080_L31E09, belonging to the ARCTIC96BD-19 clade) was the most abundant 200 m genotype. Two were Form II RuBisCO-containing targets (EB750-10B11, EB750-10A10) without phylogenetic markers but whose BLAST homology indicated relatedness to chemoautotrophic symbionts. A pelagic relative (SUP05) of these targets from Sannich Inlet was recently sequenced metagenomically, and appears to be a chemolithoautotroph that may oxidize reduced sulfur compounds, using nitrate as the terminal electron acceptor, as does its close clam-symbiont relatives (Walsh *et al.*, 2009). Although the oxygen minimum zone in Monterey Bay is significantly deeper than 200 m (generally ~700–800 m), the consistent presence of these chemoautotrophic relatives at 200 m as well as in other aerobic pelagic environments, suggests that either they may be facultatively aerobic and can chemolithoautotrophically or chemoheterotrophically thrive under oxic conditions.

In addition, three deltaproteobacterial targets were common in deep samples (with one SAR324 being *consistent* and one being *frequent*), in agreement with the previous depth preference described for this group (e.g. Wright *et al.*, 1997). These targets were also correlated to the differentiation of 200 m from 0 m and 30 m samples. Another notable *deep-consistent* target was a gammaproteobacterial genotype that clusters within a deep-sea environmental clade (that includes clones ZD0417 and DHB-2) commonly observed in 16S rRNA gene surveys from a variety of locations (López-García *et al.*, 2001). The natural history and biology of this clade remains a mystery. The genome proxy array can in this way be used to investigate the temporal and spatial dynamics of understudied but abundant organisms for which genomic fragments have been sequenced.

In addition to targeted bacteria, 3 of the 15 targeted archaea were common. Previous FISH investigations in Monterey Bay observed deep and abundant crenarchaeal populations (comprising up to 33% of the 200 m community), and euryarchaea throughout the water column at low levels (< 1%) with an increase in summer surface waters (up to 12% of the community) (Pernthaler *et al.*, 2002; Mincer *et al.*, 2007). The array signal reflected this general trend with euryarchaeal clones present in both shallow and deep samples, and the restriction of crenarchaeal targets to the deepest samples (Fig. 4), with one crenarchaeal genotype present in 57% of 200 m samples (Table 1). In addition, however, two *deep-consistent* euryarchaeal clones were among the most abundant taxa at 200 m and present in all sampling dates. This apparent inconsistency with previous observations at this site likely reflects methodological constraints of the FISH-based study, which used surface rather than deep euryarchaeal phylotypes to generate probes and thus may have missed deep genotypes. Indeed rRNA clone libraries from diverse locations have observed appreciable euryarchaeal abundances in deep waters (Massana *et al.*, 1997; López-García *et al.*, 2001; DeLong *et al.*, 2006). The array also revealed that crenarchaeal abundances paralleled those of a lower-intensity *Nitrospina* target (clone EB080_L20F04; Fig. 4), as was previously observed in a qPCR study at this site from 1997–99 (Mincer *et al.*, 2007).

Proteorhodopsin-containing taxa. Proteorhodopsin (PR) is a light-driven proton pump abundant in photic zones (Béjà *et al.*, 2000; Sabehi *et al.*, 2004; McCarren and DeLong, 2007; Rusch *et al.*, 2007) and believed to mediate photoheterotrophy in at least some of the diverse microbes that encode it (Sabehi *et al.*, 2005; Gómez-Consarnau *et al.*, 2007; Moran and Miller, 2007; Stingl *et al.*, 2007; González *et al.*, 2008). PR-containing targets accounted for 50% of the taxa (11 of 22) abundant in shallow samples (Fig. 4). Specifically, all three abundant SAR86 targets encoded PR, thought in this clade to allow photoheterotrophy (Béjà *et al.*, 2000; Sabehi *et al.*, 2004; 2005; 2007; Mou *et al.*, 2007). In addition, seven *Proteobacterial* PR-containing targets without phylogenetic markers (designated *Proteobacteria* by BLAST-based identities) were among those abundant in shallow samples.

Two of these had sufficiently inverted relative abundances at 0 m and 30 m to contribute to the differentiation of the two depths (Fig. 5A; EB000-39F01 in 0 m, and EB000-39H12 in 30 m).

In addition, three PR-containing targets (two without phylogenetic markers, and the NAC11-7 HTCC2255 genome) were among those with strong post-bloom responses. All three were also among the 10 most abundant targets in pyrosequence data, in all three sequenced post-bloom samples (circled data points in Fig. 2). This might simply reflect that these taxa were highly competitive heterotrophs under bloom conditions, with PR genes being incidental to the bloom-related phase of their lifestyle. Alternatively, PR might have allowed these taxa to persist longer than other heterotrophs as the bloom waned, as has been hypothesized for the PR-containing *Bacteroidetes* cultivar *Dokdonia* sp. MED134 (Gómez-Consarnau *et al.*, 2007). Lastly, the PR might have played a more an active role in bloom utilization, helping provide the energy for organic matter uptake and/or degradation, and allowing these heterotrophs to compete more effectively for bloom carbon.

Dynamics surrounding upwelling and bloom events. Community composition variability did not obviously correlate to Monterey Bay's three typical 'oceanographic seasons' (Fig. 4; spring/summer upwelling, fall upwelling and winter non-upwelling, as defined in, for example, Pennington and Chavez, 2000; Pennington *et al.*, 2007). However, there was substantial annual variability in the timing of the seasonal Davenport Upwelling Plume and associated upwelling events, and phytoplankton abundance and growth rates have previously been described as 'strikingly pulsed' (Pennington and Chavez, 2000). Conditions during the period sampled in this study did not follow the average seasonal breakpoints, so it is not surprising that there was little apparent correlation between sample profiles and the site's typical oceanographic seasons. Ordering the samples temporally, instead of clustering them, also did not reveal appreciable seasonal dynamics of most targets (Fig. S10). Profiling of additional years, or at higher temporal resolution, might reveal a stronger cumulative seasonal signal.

Despite the lack of a strong seasonal signal overall, the array profiles showed responses to upwelling. Following some upwelling events (as indicated by nitrate concentrations, Fig. 3), 0 m array profiles were notably intense (red starred samples in Fig. 4 and Fig. S10, and denoted by blue arrows in Fig. 3), reflecting high target abundances, and these upwelling-influenced profiles are more similar to each other than to most other 0 m or 30 m samples (as reflected in branch lengths between samples, Fig. 4). When samples are ordered temporally (Fig. S10) the seasonal nature of this response to particular spring and fall upwelling events captured by the 21 sampled dates is clear.

The phytoplankton blooms associated with upwelling are distinct between spring and fall upwelling events in Monterey Bay (Pennington *et al.*, 2007), but this difference is not reflected in the microbes profiled by the array; the post-upwelling profiles do not cluster into two distinct groups based on upwelling season. Thus, for the taxa targeted by the array, there were not recurring post-bloom communities specific to spring or fall blooms.

The post-upwelling signature in the array data was therefore at the scale of individual events rather than across seasons, and in the form of increased signal from pre-existing, common, abundant taxa rather than unique ones. The strongest target responses came from *shallow-consistent* or *-frequent* genotypes, including four NAC11-7 targets (EB080_L11F12, EB080_L43F08, EB080_L27A02 and HTCC2255) and two PR-containing alphaproteobacterial clones lacking phylomarkers (EB000-39F01, EB000-55B11). The NAC11-7 *Roseobacteria* clade is often associated with bloom and post-bloom conditions (West *et al.*, 2008, and reviewed in Buchan *et al.*, 2005), due to their common ability to degrade dimethylsulfoniopropionate, an osmolyte produced by a variety of phytoplankton. The prominent role of NAC11-7 signal at this coastal upwelling site, and their particular intensity after bloom conditions, is therefore consistent with previous observations of this clade. An additional *shallow-frequent* genotype with dramatic increase in post-bloom intensity was a representative (EB000-36A07) of the betaproteobacterial OM43 clade, which has been observed to respond to diatom blooms (Morris *et al.*, 2006), occurring in Monterey Bay during the spring/summer upwelling (Pennington *et al.*, 2007). Given that the OM43 clade appears methylotrophic (Giovannoni *et al.*, 2008), this reinforces the association between phytoplankton blooms and one-carbon compound degraders.

Responses to upwelling were also observed at 200 m. The chemical signatures of upwelling and subsequent surface bloom events were observed in patterns in nitrate, phosphate and silicate concentrations at 200 m (Fig. 3). Cold nutrient-rich water upwells through the water column; this is seen most clearly in early spring of 2004. As diatoms bloom and begin to settle through the water column, they are remineralized and may, depending on sinking and remineralization rates, produce a short-lived phosphate increase, as in mid-spring 2004. Depending on the volume of settling material, organic matter degradation may strip that water of some nutrients, which may explain the sharp drop in nitrate throughout the water column so soon after its upwelling-associated spike, concurrent with the high levels of phosphate. Remineralized nitrogen in the initial form of ammonia can be consumed before it is converted to nitrate, and existing nitrate is also

taken up by the actively degrading community. Finally, as the more recalcitrant frustule-associated component of the sinking diatomaceous organic matter becomes a higher percentage of the total available organic matter, silicate concentrations increase as silicate is remineralized. It is possible that the temporal pattern in nitrate, phosphate and silicate concentrations at 200 m, particularly evident in dramatic upwelling series in spring 2004, and the strong correlation of array profile variability to silicate and phosphate and decoupling from nitrate, represent post-diatom-bloom remineralization signatures.

A window into population heterogeneity. In addition to tracking targeted taxa, the genome proxy array design allows the tracking of close relatives of targeted strains, and through the pattern of probe hybridization can reveal population shifts over time. Population shifts were examined in two ways. First, the relative evenness of the array hybridization signal to each probe set was examined (see Rich *et al.*, 2008, and *Experimental procedures*) as a measure of the relative identity of the hybridizing genotype to the target genotype. The signal across probe sets from sporadically distributed taxa was less even than from depth-consistent taxa. It was also less even for common deep taxa compared with common shallow taxa (Fig. S11). Second, for particular targets of interest, the hybridization pattern of signal across the probe set was compared between samples. Specifically, pair-wise correlations (Pearson) of these hybridization patterns were calculated between samples. Clustering of these correlations was then used to identify samples with more or less similar probe set patterns for a given target. This process is shown for a targeted SAR86-II clone in Fig. 7, and represents complementary approaches for analysing probe signal. Averaging the signal across all probes for a given target describes the relative abundance of hybridizing genotypes, while assessing the evenness of that signal across probes (the hybridization pattern) indicates the likely genetic relatedness of hybridizing strains to the target. Then, the similarity of hybridization pattern between different samples indicates potential shifts in hybridizing populations.

As an example, all samples in which SAR86-II clone EB000-45B06 occurred (39 total; 21 samples at 0 m, 13 at 30 m and 5 at 200 m) showed similar hybridization evenness (see *Experimental procedures*). This implied similar overall identities to the targeted strain. Analysis of hybridization patterns, however, suggested the presence of four distinct populations (Fig. 7). Three of these four potential populations had cohesive occurrence patterns (occurring primarily at one depth; Fig. 7), supporting their probable existence and ecological relevance.

These results suggest the power of the genome proxy array platform to dissect fine population structure. This could be further examined by comparing the population structure of array-targeted clones to metagenomic sequence data, and will be explored in follow-up work.

Potential future use of the genome proxy array

The relative value of array versus sequencing approaches for profiling microbial communities cuts across three common research goals. (i) *Overall community profiling* ex situ: It is currently ~100-fold less expensive to repetitively characterize samples using a genome proxy array than by even the most inexpensive metagenomic methods (e.g. Illumina sequencing), and requires a fraction of the computational resources for data processing. While the array provides indirect information (hybridization patterns and intensity) on targeted genotypes and their relatives, metagenomics provides direct information about the entire community where database matches allow such inference. (ii) *Community profiling* in situ: A variety of autonomous sensors exist to perform rapid community profiling by optical (e.g. Sieracki *et al.*, 1998; Olson and Sosik, 2007; Thyssen *et al.*, 2008) or nucleic acid hybridization (e.g. Scholin *et al.*, 2001; Roman *et al.*, 2005) methods. The former discern only those few microbes with distinctive optical features. The latter currently target the 16S rRNA molecule (Preston *et al.*, 2009), although organisms with highly similar 16S sequences can have distinct ecological niches (e.g. Rocap *et al.*, 2003; Konstantinidis and Tiedje, 2005). Thus the genome proxy array approach might serve a unique methodological role on such autonomous sensors. (iii) *Population profiling*: The genome proxy array can also discern closely related populations (see above), effectively assaying both gene content and average nucleotide identity across targeted regions in related genotypes. While metagenomic data can provide population inferences, these have been limited to cases where assemblies are possible (e.g. low-diversity environments, Tyson *et al.*, 2004, or dominant taxa in more complex communities, Venter *et al.*, 2004), or to small sequence reads that represent ~40-fold less of the genome than the genome proxy array. Thus, for now, the genome proxy array retains utility as an *ex situ* community profiling tool, and complements sequencing for applications of *in situ* profiling and population tracking.

Conclusions

Exploration of the array profiles and the underlying causes of their variability allowed a cost-effective understanding of target natural history, and of community dynamics over time. Thus far, we tracked the genotype abundances of 268 target taxa through 57 samples collected over 4 years in Monterey Bay, at three oceano-

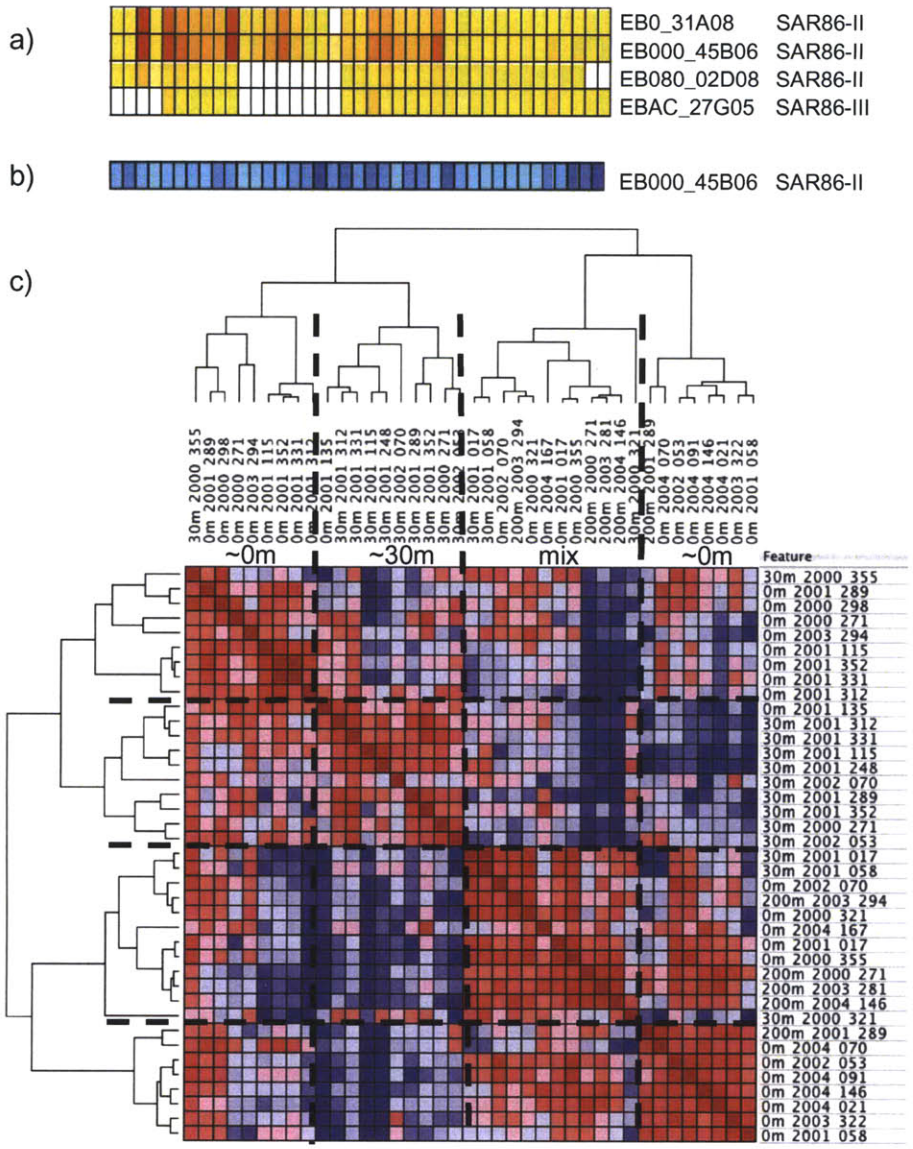


Fig. 7. Revealing population heterogeneity by the genome proxy array: complementary probe set analyses moving from overall target abundance to strain and population information.
A. Mean target intensity for SAR86 target strains present in Monterey Bay samples (as in Fig. 4A). EB000_45B06 is ubiquitous in shallow samples.
B. Relative evenness of hybridization signal across the SAR86-II target EB000_45B06 target probe set (as Tukey biweight-over-mean value; see *Experimental procedures*). By this index alone, subpopulations are not strongly evident.
C. Pair-wise Pearson correlations of the signal pattern across the EB000_45B06 probe set, between every sample in which it occurred. Samples are clustered based on similarity of probe set pattern (assessed by Pearson correlation). Four major clusters of samples are present, delineated by black dashed lines, evident in both the clustering patterns and in the matrix diagonal. Red indicates high Pearson correlation, white is intermediate, blue is low.

graphically distinct depths (Fig. 3). While the targets were distributed across known marine microbial diversity and had diverse geographic origins, 95 targeted taxa were present in at least one sample, and 31 were present in > 50% of samples. Most taxa showed differential distribution with depth (Fig. 4). Highly abundant shallow taxa included representatives of the SAR86, SAR116, SAR11 and *Roseobacter* clades. Notably, the majority of abundant shallow taxa contained the proteorhodopsin gene. Highly abundant deep taxa included representatives of marine pelagic euryarchaea, deltaproteobacteria (including the SAR324 clade), and relatives of invertebrate

chemoautotrophic symbionts. All 200 m samples clustered together to the exclusion of 0 m and 30 m samples, although there was no clear clustering of each of the shallower depths. No clustering-based correlation of sample profile to oceanographic season was seen, but overall profile intensity 'blooms' were observed in profiles after episodic upwelling events, and possible post-bloom remineralization events were indicated in several 200 m samples. Finally, the array suggested that some targets were present as multiple distinct populations over time and space; these population dynamics suggest new directions for future research on microbial population dynamics.

Experimental procedures

Sampling and DNA extractions

Samples were collected from Station M1 (36.747°N, 122.022°W) in Monterey Bay at approximately monthly intervals, with several longer gaps, between JD271 in 2000 and JD167 in 2004. Two litres of seawater from each of eight depths (0, 20, 30, 40, 80, 100, 150 and 200 m) was filtered through a 45 mm GF-A 1.6-μm-pore pre-filter (Whatman) and concentrated onto a 25 mm Supor-200 0.2-μm-pore filter (Pall Corp, Ann Arbor, MI), using a MasterFlex peristaltic pump system (Cole-Parmer Instrument Company, Vernon Hills, IL) at ≤ 15 psi. Filters were stored dry in 2 ml screw-cap tubes, immediately placed in a −20°C freezer shipboard, and transferred on ice to a −80°C freezer upon landfall.

DNA was extracted from all 0 m and 200 m filters available from 2000 JD271 through 2004 JD167, and all 30 m samples available from 2000 JD271 through 2002 JD070. In this location, 0 m is in the photic zone, 30 m is generally below the mixed layer, and 200 m is below the photic zone. All MB DNA extractions were performed simultaneously in 96-well format to minimize extraction variability, as in Rich and colleagues (2008). Briefly, cell lysis was performed by incubating each filter with 242 ml lysis buffer (lysis buffer: 40 mM EDTA, 50 mM Tris pH 8.3, 0.73 M sucrose, 1.15 mg ml^{-1} lysozyme, 200 mg ml^{-1} RNase, 0.2 mm filter-sterilized) in a microcentrifuge tube at 37°C for 30 min, rotating. Protein degradation was accomplished by adding SDS to 1%, and 13.5 ml of Proteinase K solution (10 mg ml^{-1} in 40 mM EDTA, 50 mM Tris pH 8.3, 0.73 M sucrose), and incubating overnight at 55°C, rotating. DNA was then extracted with the DNeasy 96 Tissue kit (Qiagen, Valencia, CA), using modifications of the manufacturer's protocol. Each tube was vortexed with 300 ml of Buffer AL and incubated at 70°C for 10 min, then vortexed with 300 ml of 99% ethanol and pipetted onto a 96-well spin plate. The plate was sealed with an airpore sheet (supplied with kit) and spun at 40°C, 4612 *g* in a Sorvall Legend RT centrifuge (Kendro Laboratory Products, Newtown, CT). After a 10 min spin 500 ml of Buffer AW1 was added to each well, the plate was re-sealed and spun 5 min, then 500 ml of Buffer AW2 was added to each well, and the plate was re-sealed and spun 5 min. Columns were then incubated for 15 min at 70°C atop a new rack of elution microtubes RS (supplied with kit). DNA was eluted with 2 × 200 ml of Buffer AE preheated to 70°C, incubated 1 min and spun 2 min. Finally, DNA was concentrated by Excela-Pure 96-well PCR purification kits (Edge BioSystems, Gaithersburg, MD), following the manufacturer's protocol. DNA was rinsed with 100 ml of nuclease-free water, resuspended in 20 ml of dilute TE (1 mM Tris pH 8, 0.1 mM EDTA pH 8), and transferred to a clean 96-well plate. Extracted DNAs were quantified spectrophotometrically (Nanodrop, Thermo Scientific) and stored at −80°C until use. Yields averaged ~470 ng per litre of seawater for 200 m samples (range 177–903 ng) and ~1460 ng per litre of seawater for 0 m and 30 m samples (range 484–3804 ng).

Oceanographic data

Oceanographic data were kindly provided by Reiko Michisaki and Francisco Chavez of the Biological Oceanography Group at the Monterey Bay Aquarium Research Institute, who collected and processed it as part of the Monterey Bay time-series programme. Measurement methods were described in Asanuma and colleagues (1999). Nutrient (nitrate, nitrite, silicate and phosphate) data used for correlation analyses are in Table S4, and additional plots can be accessed at http://www.mbari.org/bog/.

Arrays design, hybridization and data processing

The expanded genome proxy array was designed as in Rich and colleagues (2008). Briefly, each genotype was targeted using suites of ~20 70-mer oligonucleotide probes designed using the program ArrayOligoSelector (Zhu *et al.*, 2003). Probes had approximately the same %GC (40%) and were distributed across the target genome or genome fragment, with no more than one probe per gene and avoiding 16S and 23S rRNA genes. The array included positive and negative control probes designed using the same method, to *Halobacterium salinarum* NRC-1 and a random genome sequence respectively.

The expanded array had a broader scope than the prototype of Rich and colleagues (2008) (268 target genotypes, as opposed to the prototype's 14) and included a co-spot oligo for spot alignment and gridding purposes (using the 'alien' oligo sequence of Urisman *et al.*, 2005). The targets were selected from fully sequenced marine microbial genomes, publicly available marine-derived BAC and fosmid clone sequences, and fully sequenced clones from the lab's Monterey Bay and Hawaii environmental BAC- and fosmid-based genomic libraries. Targeted genotypes are detailed in Table S1, summarized in Table S2, and presented in a schematic phylogenetic overview in Fig. 1. Previously unpublished sequences used for array design were submitted to GenBank under Accession No. GU474833–GU474949.

Hybridizations were performed as in Rich and colleagues (2008), by labelling randomly amplified sample DNA with a single fluorophore (Cy3) for hybridization. The following modifications were made to the Rich and colleagues (2008), hybridization method: Round A, B and C amplification reactions were performed in 96-well plates for higher throughput,

and cleaned through ExcelaPure 96-well plates (Edge Biosystems, Gaithersburg). They were washed twice with 300 µl of TE, dried down and resuspended directly in 0.1 M $NaHCO_3$ for the labelling reactions. Approximately 1 pmol of Cy5-labelled co-spot complement oligo was added to each hybridization for spot localization purposes (modified from Urisman *et al.*, 2005). For each sample, at least three replicate arrays were hybridized. (As arrays constructed in-house, some did not produce high-quality data due to significant surface peeling of the poly-lysine coating during hybridization or excessive background fluorescence; ~20% of arrays were discarded and additional arrays were hybridized.)

Data were pre-processed as in Rich and colleagues (2008), with minor modifications. Briefly, poorly performing arrays, defined as those with less than half the positive control probes brighter than the standard deviation of the negative control probes, were removed from further analysis. Within each remaining array, bad spots (those with areas of poly-L-lysine peeling or excessive background fluorescence) were manually flagged and removed from further analysis. Background-subtracted spot intensities were negative-control-subtracted and normalized to each array's mean positive control value, then replicate spots of a given probe were pooled across arrays and the median was taken as the value for that probe.

Finally, the signal for each targeted genotype was calculated. To be considered present, at least 40% of its probes were required to be above the standard deviation of the negative control probe set (rather than above twice the mean negative control value, as in Rich *et al.*, 2008), or the targeted genotype was considered 'absent' and its value set to zero. This was done to remove erroneous target abundances due to uninformative single-gene cross-hybridizations. For targets that passed this thresh-holding step, the mean or Tukey biweight (TBW) across each probe set was taken, as in Rich and colleagues (2008). We did not examine which probes for each organism showed signal, since probes were not designed to distinguish particular genes; i.e. no alignments were used to target conserved or variable parts of given genes, but instead the probe was chosen purely on hybridization characteristics.

Array platform design and hybridization data were deposited in the Gene Expression Omnibus, under platform Accession No. GPL10357 and samples GSM537253-310.

Data analyses

Clustering analyses of sample hybridization data were performed in GenePattern (Reich *et al.*, 2006), using hierarchical clustering (Eisen *et al.*, 1998) by Pearson correlations for both rows and columns, using pair-wise complete linkage, and without row or column centring. Principal component analysis (PCA) was performed both in GenePattern and in R using the prcomp function. Canonical discriminant analyses (CDA) were performed in R with the candisc function. In order to keep the number of variables less than the number of responses (i.e. samples), CDA was performed using the top 28 principal components instead of all detected organisms. Correlations were calculated between environmental parameters or organism abundances and each plotted principal component or canonical discriminant axis. The relative values of the correlations were represented as vectors on the analysis graphs.

Array-versus-pyrosequencing comparisons

Three 0 m samples were chosen for parallel pyrosequencing and array hybridization, based on their DNA yields. Approximately 3 µg each of samples 2000 JD298, 2001 JD115 and 2001 JD135 were sequenced at the Schuster Lab pyrosequencing facility (Pennsylvania State University) on a GS-FLX DNA sequencer (454 Life Sciences, Brandford, CT).

Sequence clean-up. To remove poor-quality pyrosequences, the length distribution of the raw reads for each sample was plotted. From the empirical cumulative density function (ecdf) plot, the lower and upper boundary lengths were estimated so that 95% of the read lengths fell between the boundaries (which varied for each sample: 71 and 305 bp for 2000JD298, 65 and 255 bp for 2001JD115, and 65 and 303 bp for 2001JD135). The outlying 5% of the reads were removed. Reads with more than one 'N' were also removed. This two-step process removed approximately 5.5% of the reads overall; for 2000JD298, 23 917 out of 419 684 reads (5.7%) were discarded, for 2001JD115, 19 822 out of 365 472 reads (5.4%) were discarded, and for 2001JD135, 22 887 out of 414 861 reads (5.5%) were discarded.

BLASTN parameters. To identify BLASTN parameters that would give the closest *in silico* similarity to the array's range of cross-hybridization, we used the genomes of *Prochlorococcus* MED4, MIT9515 and MIT9312, whose relative hybridization strength to the array's strain MED4 probes was measured previously (Rich *et al.*, 2008). The genomes were fragmented *in silico* into overlapping (tiled) 100 bp fragments using a perl script (kindly provided by G. Tyson), and each set of fragments was BLASTed against the MED4 genome to compare self-self (MED4 to MED4, 100% identity), MIT9515-versus-MED4 (86% average genomic identity, calculated as in Konstantinidis and Tiedje, 2005), and MIT9312-versus-MED4 results (78.5% average genomic identity). A variety of command-line BLASTN parameters were tested for similarity of results to those of the array: (i) X150 q-1 r1 W7 FF, (ii) X30 q-3 r1 W7 FF, (iii) X30 q-5 r1 W7 FF, (iv) X30 q-5 r2 W7 FF and (v) X30 q-7 r2 W7 FF. The first parameter set (X150 q-1 r1 W7 FF) yielded the best separation of the distribution of MED4–MED4 hits from MED4–MIT9515 and MED4–MIT9312 hits, and was subsequently used in downstream analyses.

Parsing parameters. BLASTN hits to a given target were parsed by bit score. However, because pyrosequencing reads range in lengths, and read length effects bit score, we investigated the correlation between read length and bit score for MIT9515 fragments versus MED4, and for MIT9312 fragments versus MED4. In addition to tiled 100 bp fragments, tiled 50 bp, 75 bp and 125 bp fragments were also generated. Linear equations for bit score (y-axis) versus read length (x-axis) were determined. The MED4–MIT9312 slope was smaller than that of MED4–MIT9515, due to the lower average identity involved at any given read length. Since cross-hybridization at or above the MIT9515–MED4 level of

identity dominates the signal of the microarray (Rich *et al.*, 2008), the equation for that comparison was used to adjust the bit score to the read length for each individual read.

Monterey Bay pyrosequencing versus array comparison. Using the BLASTN parameters and parsing criteria optimized above, the reads from each pyrosequenced Monterey Bay sample were BLASTed against all 268 genomes and genome fragments to which the array was targeted. Reads were assigned to (i.e. recruited to) one or more array targets, proportional to their bitscore, to mimic the cross-hybridization permitted by the array. Thus, if one read matched three targets using the criteria outlined above, then it would be assigned to the first of those targets as 1 * [bitscore1/(bitscore1 + bitscore2 + bitscore3)], to the second as 1 * [bitscore2/(bitscore1 + bitscore2 + bitscore3)], etc. The read-based recruitment abundance of each array target was then normalized to the length of the target query, and to the database size. For each of the three samples, the pyrosequence-based abundances of each genotype were then compared with the array-based abundances. Despite a full plate of sequencing per sample, recruitment of reads to each target was insufficient to screen presence/absence based on the signal evenness across each target, a standard step in the array data analysis pipeline. Therefore, unthresholded array data without the evenness filter (i.e. the signal for each organism before requiring at least 40% of its probes to be above the described threshold) were compared with pyrosequencing data for each target genotype.

Acknowledgements

We gratefully acknowledge the captain and crew of the R.V. *Point Lobos* for expert assistance at sea and Drs Christina Preston and Lynne Christianson for sample collection over 4 years. We also thank Francisco Chavez and Reiko Michisaki of the MBARI Biological Oceanography Group for the corresponding chemical oceanographic time-series data. Lastly, we thank Matt Sullivan and three anonymous reviewers for helpful comments on the manuscript. This work was supported by grants to E.F.D. from the Gordon and Betty Moore Foundation, a National Science Foundation award EF 0424599 (C-MORE), NSF Microbial Observatory Award MCB-0348001, and the Office of Science (BER) US Department of Energy.

References

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., and Lipman, D.J. (1990) Basic local alignment search tool. *J Mol Biol* **215:** 403–410.

Arrigo, K.R. (2005) Marine microorganisms and global nutrient cycles. *Nature* **437:** 349–355.

Asanuma, H., Rago, T.A., Collins, C.A., Chavez, F.P., and Castro, C.G. (1999) *Changes in the Hydrography of Central California Waters Associated with the 1997–1998.* Monterey, CA, USA: Naval Postgraduate School.

Bano, N., and Hollibaugh, J.T. (2002) Phylogenetic composition of bacterioplankton assemblages from the Arctic Ocean. *Appl Environ Microbiol* **68:** 505–518.

Béjà, O., Aravind, L., Koonin, E.V., Suzuki, M.T., Hadd, A., Nguyen, L.P., *et al.* (2000) Bacterial rhodopsin: evidence for a new type of phototrophy in the sea. *Science* **289:** 1902–1906.

Buchan, A., Gonzalez, J.M., and Moran, M.A. (2005) Overview of the marine roseobacter lineage. *Appl Environ Microbiol* **71:** 5665–5677.

Dalsgaard, T., Canfield, D.E., Petersen, J., Thamdrup, B., and Acuna-Gonzalez, J. (2003) N_2 production by the anammox reaction in the anoxic water column of Golfo Dulce, Costa Rica. *Nature* **422:** 606–608.

DeLong, E.F., Preston, C.M., Mincer, T., Rich, V., Hallam, S.J., Frigaard, N.U., *et al.* (2006) Community genomics among stratified microbial assemblages in the ocean's interior. *Science* **311:** 496–503.

Dinsdale, E.A., Pantos, O., Smriga, S., Edwards, R.A., Angly, F., Wegley, L., *et al.* (2008) Microbial ecology of four coral atolls in the Northern Line Islands. *PLoS ONE* **3:** e1584.

Eilers, H., Pernthaler, J., Glockner, F.O., and Amann, R. (2000) Culturability and *in situ* abundance of pelagic bacteria from the North Sea. *Appl Environ Microbiol* **66:** 3044–3051.

Eisen, M.B., Spellman, P.T., Brown, P.O., and Botstein, D. (1998) Cluster analysis and display of genome-wide expression patterns. *Proc Natl Acad Sci USA* **95:** 14863–14868.

Field, K.G., Gordon, D., Wright, T., Rappé, M., Urback, E., Vergin, K., and Giovannoni, S.J. (1997) Diversity and depth-specific distribution of SAR11 cluster rRNA genes from marine planktonic bacteria. *Appl Environ Microbiol* **63:** 63–70.

Fuhrman, J.A., McCallum, K., and Davis, A.A. (1992) Novel major archaebacterial group from marine plankton. *Nature* **356:** 148–149.

Fuhrman, J.A., Hewson, I., Schwalbach, M.S., Steele, J.A., Brown, M.V., and Naeem, S. (2006) Annually reoccurring bacterial communities are predictable from ocean conditions. *Proc Natl Acad Sci USA* **103:** 13104–13109.

Giovannoni, S.J., and Rappé, M.S. (2000) Evolution, diversity and molecular ecology of marine prokaryotes. In *Microbial Ecology of the Oceans.* Kirchman, D.L. (ed.). New York, NY, USA: Wiley and Sons, pp. 47–84.

Giovannoni, S.J., Hayakawa, D.H., Tripp, H.J., Stingl, U., Givan, S.A., Cho, J.-C., *et al.* (2008) The small genome of an abundant coastal ocean methylotrophs. *Environ Microbiol* **10:** 1771–1782.

Gómez-Consarnau, L., González, J.M., Coll-Lladó, M., Gourdon, P., Pascher, T., Neutze, R., *et al.* (2007) Light stimulates growth of proteorhodopsin-containing marine Flavobacteria. *Nature* **445:** 210–213.

González, J.M., Fernández-Gómez, B., Fernàndez-Guerra, A., Gómez-Consarnau, L., Sánchez, O., Coll-Lladó, M., *et al.* (2008) Genome analysis of the proteorhodopsin-containing marine bacterium *Polaribacter* sp. MED152 (Flavobacteria). *Proc Natl Acad Sci USA* **105:** 8724–8729.

Howard, E.C., Henriksen, J.R., Buchan, A., Reisch, C.R., Burgmann, H., Welsh, R., *et al.* (2006) Bacterial taxa that limit sulfur flux from the ocean. *Science* **314:** 649–652.

Karl, D.M. (1999) A sea of change: biogeochemical variability in the North Pacific Subtropical Gyre. *Ecosystems* **2:** 181–214.

Karl, D.M. (2007) Microbial oceanography: paradigms, processes and promise. *Nat Rev Microbiol* **5:** 759–769.

Karner, M.B., DeLong, E.F., and Karl, D.M. (2001) Archaeal dominance in the mesopelagic zone of the Pacific Ocean. *Nature* **409:** 507–510.

Kennedy, J., Marchesi, J., and Dobson, A. (2007) Metagenomic approaches to exploit the biotechnological potential of the microbial consortia of marine sponges. *Appl Microbiol Biotechnol* **75:** 11–20.

Klepac-Ceraj, V. (2004) Diversity and phylogenetic structure of two complex marine microbial communities. Thesis. Cambridge, MA, USA: Department of Civil and Environmental Engineering, Massachusetts Institute of Technology.

Kolber, Z.S., Van Dover, C.L., Niederman, R.A., and Falkowski, P.G. (2000) Bacterial photosynthesis in surface waters of the open ocean. *Nature* **407:** 177–179.

Konstantinidis, K.T., and Tiedje, J.M. (2005) Genomic insights that advance the species definition for prokaryotes. *Proc Natl Acad Sci USA* **102:** 2567–2572.

Kuypers, M.M.M., Sliekers, A.O., Lavik, G., Schmid, M., Jorgensen, B.B., Kuenen, J.G., *et al.* (2003) Anaerobic ammonium oxidation by anammox bacteria in the Black Sea. *Nature* **422:** 608–611.

López-García, P., López-López, A., Moreira, D., and Rodríguez-Valera, F. (2001) Diversity of free-living prokaryotes from a deep-sea site at the Antarctic Polar Front. *FEMS Microbiol Ecol* **36:** 193–202.

Ludwig, W., Strunk, O., Westram, R., Richter, L., Meier, H., Yadhukumar, *et al.* (2004) ARB: a software environment for sequence data. *Nucleic Acids Res* **32:** 1363–1371.

McCarren, J., and DeLong, E.F. (2007) Proteorhodopsin photosystem gene clusters exhibit co-evolutionary trends and shared ancestry among diverse marine microbial phyla. *Environ Microbiol* **9:** 846–858.

Marhaver, K.L., Edwards, R.A., and Rohwer, F. (2008) Viral communities associated with healthy and bleaching corals. *Environ Microbiol* **10:** 2277–2286.

Massana, R., Murray, A.E., Preston, C.M., and DeLong, E.F. (1997) Vertical distribution and phylogenetic characterization of marine planktonic Archaea in the Santa Barbara Channel. *Appl Environ Microbiol* **63:** 50–56.

Mincer, T.J., Church, M.J., Taylor, L.T., Preston, C., Karl, D.M., and DeLong, E.F. (2007) Quantitative distribution of presumptive archaeal and bacterial nitrifiers in Monterey Bay and the North Pacific Subtropical Gyre. *Environ Microbiol* **9:** 1162–1175.

Moran, M.A., and Miller, W.L. (2007) Resourceful heterotrophs make the most of light in the coastal ocean. *Nat Rev Microbiol* **5:** 792.

Morris, R.M., Rappé, M.S., Connon, S.A., Vergin, K.L., Siebold, W.A., Carlson, C.A., and Giovannoni, S.J. (2002) SAR11 clade dominates ocean surface bacterioplankton communities. *Nature* **420:** 806–810.

Morris, R.M., Rappe, M.S., Urbach, E., Connon, S.A., and Giovannoni, S.J. (2004) Prevalence of the *Chloroflexi*-related SAR202 bacterioplankton cluster throughout the mesopelagic zone and deep ocean. *Appl Environ Microbiol* **70:** 2836–2842.

Morris, R.M., Longnecker, K., and Giovannoni, S.J. (2006) *Pirellula* and OM43 are among the dominant lineages identified in an Oregon coast diatom bloom. *Environ Microbiol* **8:** 1361–1370.

Morris, R., Vergin, K., Cho, J.-C., Rappé, M., Carlson, C., and Giovannoni, S. (2005) Temporal and spatial response of bacterioplankton lineages to annual convective overturn at the Bermuda Atlantic Time-series study site. *Limnol Oceanogr* **50:** 1687–1696.

Mou, X., Hodson, R.E., and Moran, M.A. (2007) Bacterioplankton assemblages transforming dissolved organic compounds in coastal seawater. *Environ Microbiol* **9:** 2025–2037.

Mou, X., Sun, S., Edwards, R.A., Hodson, R.E., and Moran, M.A. (2008) Bacterial carbon processing by generalist species in the coastal ocean. *Nature* **451:** 708–711.

Mullins, T.D., Britschgi, T.B., Krest, R.L., and Giovannoni, S.J. (1995) Genetic comparisons reveal the same unknown bacterial lineages in Atlantic and Pacific bacterioplankton communities. *Limnol Oceanogr* **40:** 148–158.

Neufeld, J.D., Chen, Y., Dumont, M.G., and Murrell, J.C. (2008) Marine methylotrophs revealed by stable-isotope probing, multiple displacement amplification and metagenomics. *Environ Microbiol* **10:** 1526–1535.

O'Mullan, G.D., and Ward, B.B. (2005) Relationship of temporal and spatial variabilities of ammonia-oxidizing bacteria to nitrification rates in Monterey Bay, California. *Appl Environ Microbiol* **71:** 697–705.

Olson, R.J., and Sosik, H.M. (2007) A submersible imaging-in-flow instrument to analyze nano- and microplankton: imaging FlowCytobot. *Limnol Oceanogr: Methods* **5:** 195–203.

Partensky, F., Blanchot, J., and Vaulot, D. (1999) Differential distribution and ecology of *Prochlorococcus* and *Synechococcus* in oceanic waters: a review. In *Marine Cyanobacteria and Related Organisms.* Charpy, L., and Larkum, H. (eds). Monaco: Musée océanographique. Bulletin de l'Institut Océanographique (Monaco) NS**19**, pp. 431–449.

Pennington, J.T., and Chavez, F.P. (2000) Seasonal fluctuations of temperature, salinity, nitrate, chlorophyll and primary production at station H3/M1 over 1989–1996 in Monterey Bay, California. *Deep Sea Res Part II Top Stud Oceanogr* **47:** 947–973.

Pennington, J.T., Michisaki, R., Johnston, D., and Chavez, F.P. (2007) *Ocean Observing in the Monterey Bay National Marine Sanctuary: CalCOFI and the MBARI Time Series.* Monterey, CA, USA: The Sanctuary Integrated Monitoring Network (SIMoN), Monterey Bay Sanctuary Foundation, and Monterey Bay National Marine Sanctuary.

Pernthaler, A., Preston, C.M., Pernthaler, J., DeLong, E.F., and Amann, R. (2002) Comparison of fluorescently labeled oligonucleotide and polynucleotide probes for the detection of pelagic marine bacteria and archaea. *Appl Environ Microbiol* **68:** 661–667.

Preston, C.M., Marin, R., Jensen, S.D., Feldman, J., Birch, J.M., Massion, E.I., *et al.* (2009) Near real-time autonomous detection of marine bacterioplankton on a coastal mooring in Monterey Bay, California, using rRNA-targeted DNA probes. *Environ Microbiol* **11:** 1168–1180.

Pruesse, E., Quast, C., Knittel, K., Fuchs, B., Ludwig, W., Peplies, J., and Glöckner, F.O. (2007) SILVA: a comprehensive online resource for quality checked and aligned

ribosomal RNA sequence data compatible with ARB. *Nucleic Acids Res* **35:** 7188–7196.

Rappé, M.S., Vergin, K., and Giovannoni, S.J. (2000) Phylogenetic comparisons of a coastal bacterioplankton community with its counterparts in open ocean and freshwater systems. *FEMS Microbiol Ecol* **33:** 219–232.

Reich, M.L.T., Gould, J., Lerner, J., Tamayo, P., and Mesirov, J.P. (2006) GenePattern 2.0. *Nat Genet* **38:** 500–501.

Rich, V.I., Konstantinidis, K., and DeLong, E.F. (2008) Design and testing of 'genome-proxy' microarrays to profile marine microbial communities. *Environ Microbiol* **10:** 506–521.

Rocap, G., Larimer, F.W., Lamerdin, J., Malfatti, S., Chain, P., Ahlgren, N.A., *et al.* (2003) Genome divergence in two *Prochlorococcus* ecotypes reflects oceanic niche differentiation. *Nature* **424:** 1042–1047.

Roman, B., Scholin, C., Jensen, S., Marin, R., Massion, E., and Feldman, J. (2005) The 2nd generation environmental sample processor: evolution of a robotic underwater biochemical laboratory. In *OCEANS 2005 MTS/IEEE Conference.* Washington, DC, USA: Marine Technology Society.

Rusch, D.B., Halpern, A.L., Sutton, G., Heidelberg, K.B., Williamson, S., Yooseph, S., *et al.* (2007) The Sorcerer II Global Ocean Sampling expedition: northwest Atlantic through eastern tropical Pacific. *PLoS Biol* **5:** e77.

Sabehi, G., Béjà, O., Suzuki, M.T., Preston, C.M., and DeLong, E.F. (2004) Different SAR86 subgroups harbour divergent proteorhodopsins. *Environ Microbiol* **6:** 903–910.

Sabehi, G., Loy, A., Jung, K.-H., Partha, R., Spudich, J.L., Isaacson, T., *et al.* (2005) New insights into metabolic properties of marine bacteria encoding proteorhodopsins. *PLoS Biol* **3:** e273.

Sabehi, G., Kirkup, B.C., Rozenberg, M., Stambler, N., Polz, M.F., and Béjà, O. (2007) Adaptation and spectral tuning in divergent marine proteorhodopsins from the eastern Mediterranean and the Sargasso Seas. *ISME J* **1:** 48–55.

Scholin, C.A., Massion, E.I., Wright, D., Cline, D., Mellinger, E., and Brown, M. (2001) Aquatic Autosampler Device. US patent 6187530.

Sieracki, C.K, Sieracki, M.E., and Yentsch, C.S. (1998) An imaging-in-flow system for automated analysis of marine microplankton. *Mar Ecol Prog Ser* **168:** 285–296.

Stevens, H., and Ulloa, O. (2008) Bacterial diversity in the oxygen minimum zone of the eastern tropical South Pacific. *Environ Microbiol* **10:** 1244–1259.

Stingl, U., Tripp, H.J., and Giovannoni, S.J. (2007) Improvements of high-throughput culturing yielded novel SAR11 strains and other abundant marine bacteria from the Oregon coast and the Bermuda Atlantic Time Series study site. *ISME J* **1:** 361–371.

Suzuki, M.T., Béjà, O., Taylor, L.T., and Delong, E.F. (2001a) Phylogenetic analysis of ribosomal RNA operons from uncultivated coastal marine bacterioplankton. *Environ Microbiol* **3:** 323–331.

Suzuki, M.T., Preston, C.M., Chavez, F.P., and DeLong, E.F. (2001b) Quantitative mapping of bacterioplankton populations in seawater: field tests across an upwelling plume in Monterey Bay. *Aquat Microb Ecol* **24:** 117–127.

Suzuki, M.T., Preston, C.M., Béjà, O., Torre, J.R., Steward, G.F., and DeLong, E.F. (2004) Phylogenetic screening of ribosomal RNA gene-containing clones in bacterial artificial chromosome (BAC) libraries from different depths in Monterey Bay. *Microb Ecol* **48:** 473–488.

Thyssen, M., Tarran, G.A., Zubkov, M.V., Holland, R.J., Gregori, G., Burkill, P.H., and Denis, M. (2008) The emergence of automated high-frequency flow cytometry: revealing temporal and spatial phytoplankton variability. *J Plankton Res* **30:** 333–343.

Treusch, A.H., Vergin, K.L., Finlay, L.A., Donatz, M.G., Burton, R.M., Carlson, C.A., and Giovannoni, S.J. (2009) Seasonality and vertical structure of microbial communities in an ocean gyre. *ISME J* **3:** 1148–1163.

Tringe, S.G., von Mering, C., Kobayashi, A., Salamov, A.A., Chen, K., Chang, H.W., *et al.* (2005) Comparative metagenomics of microbial communities. *Science* **308:** 554–557.

Tyson, G.W., Chapman, J., Hugenholtz, P., Allen, E.E., Ram, R.J., Richardson, P.M., *et al.* (2004) Community structure and metabolism through reconstruction of microbial genomes from the environment. *Nature* **428:** 37–43.

Urisman, A., Fischer, K.F., Chiu, C.Y., Kistler, A.L., Beck, S., Wang, D., and DeRisi, J.L. (2005) E-Predict: a computational strategy for species identification based on observed DNA microarray hybridization patterns. *Genome Biol* **6:** R78.

Venter, J.C., Remington, K., Heidelberg, J.F., Halpern, A.L., Rusch, D., Eisen, J.A., *et al.* (2004) Environmental genome shotgun sequencing of the Sargasso Sea. *Science* **304:** 66–74.

Walsh, D.A., Zaikova, E., Howes, C.G., Song, Y.C, Wright, J.J., Tringe, S.G., *et al.* (2009) Metagenome of a versatile chemolithoautotroph from expanding oceanic dead zones. *Science* **326:** 578–582.

Ward, B.B. (2005) Temporal variability in nitrification rates and related biogeochemical factors in Monterey Bay, California, USA. *Mar Ecol Prog Ser* **292:** 97–109.

Waterbury, J.B., Watson, S.W., Guillard, R.R.L., and Brand, L.E. (1979) Wide-spread occurrence of a unicellular, marine planktonic, cyanobacterium. *Nature* **277:** 293–294.

Waterbury, J.B., Watson, S.W., Valois, F.W., and Franks, D.G. (1986) Biological and ecological characterization of the marine unicellular cyanobacterium *Synechococcus. Can Bull Fish Aquat Sci* **214:** 71–120.

Wegley, L., Edwards, R., Rodriguez-Brito, B., Liu, H., and Rohwer, F. (2007) Metagenomic analysis of the microbial community associated with the coral *Porites astreoides. Environ Microbiol* **9:** 2707–2719.

West, N.J., Obernosterer, I., Zemb, O., and Lebaron, P. (2008) Major differences of bacterial diversity and activity inside and outside of a natural iron-fertilized phytoplankton bloom in the Southern Ocean. *Environ Microbiol* **10:** 738–756.

Wilhelm, L.J., Tripp, H.J., Givan, S.A., Smith, D.P., and Giovannoni, S.J. (2007) Natural variation in SAR11 marine bacterioplankton genomes inferred from metagenomic data. *Biol Direct* **2:** 27.

Wright, T.D., Vergin, K.L., Boyd, P.W., and Giovannoni, S.J. (1997) A novel delta-subdivision proteobacterial lineage from the lower ocean surface layer. *Appl Environ Microbiol* **63:** 1441–1448.

Yooseph, S., Sutton, G., Rusch, D.B., Halpern, A.L., Williamson, S.J., Remington, K., *et al.* (2007) The Sorcerer II

Global Ocean Sampling expedition: expanding the universe of protein families. *PLoS Biol* **5:** e16.
Zhu, J., Bozdech, Z., and DeRisi, J. (2003) *Array Oligo Selector* [WWW document]. URL http://arrayoligosel.sourceforge.net/.
Zubkov, M.V., Fuchs, B.M., Archer, S.D., Kiene, R.P., Amann, R., and Burkill, P.H. (2002) Rapid turnover of dissolved DMS and DMSP by defined bacterioplankton communities in the stratified euphotic zone of the North Sea. *Deep Sea Res Part II Top Stud Oceanogr* **49:** 3017–3038.

Supporting information

Additional Supporting Information may be found in the online version of this article:

Figs S1–S5. Phylogenetic trees illustrating the relationship of SSU rRNA gene sequences from genomes and uncultivated clones represented on the genome-proxy microarray (blue) and their close relatives (black) as 'landmarks'. Support for dendrogram topologies is indicated by bootstrap values at nodes determined by the maximum likelihood method (only values > 50 are shown). The outgroups used were *Methanomethylovorans victoriae* strain TM (AJ276437) for the bacterial dendrograms, and *Myxococcus xanthus* strain UCDaV1 (AY724797) for the archaeal dendrogram. *The publicly available SSU rDNA sequence for the *Roseobacter*-like alphaproteobacterial clone HTCC2255 (AATR01000062) is from a *Gammaproteobacterium*, known to have contaminated the HTCC2255 culture (http://www.roseobase.org/roseo/htcc2255.html). **S1.** *Gamma-* and *Betaproteobacteria*. **S2.** *Alphaproteobacteria*. **S3.** *Deltaproteobacteria* and *Spirochaetes*. **S4.** Other *Bacteria*. **S5.** *Archaea*.
Fig. S2. Alphaproteobacterial array targets (blue) and their close 'landmark' relatives (black).
Fig. S3. Deltaproteobacterial and Spirochaete array targets (blue) and their close 'landmark' relatives (black).
Fig. S4. Other bacterial array targets (blue) and their close 'landmark' relatives (black).
Fig. S5. Archaeal array targets (blue) and their close 'landmark' relatives (black).
Fig. S6. Origin of array targets and their relative array-based occurrences in Monterey Bay and Hawaii samples.
A. Derivation of array targets, either as environmental genome fragments from Hawaii (blue), Monterey (green), other marine sites (beige), or from marine microbial genomes (black). The number of targets in each category is indicated.
B. The proportional abundance of each target type in 57 Monterey Bay samples, measured as the relative proportion of total array signal across all samples hybridized.
Fig. S7. Mixed layer depth (MLD) over the sampling period, with hybridized samples indicated. MLD was calculated as the first depth (≥ 10 m) with > 0.1°C difference from the previous meter (per MBARI BOG group, R. Michisaki, pers. comm.). *X*-axis indicates sampling date in continuous numbered days since 1 January 2000, and *y*-axis indicates depth. Dashed red line highlights 30 m depth. Trendline shows moving average of MLD with period of 2. The MLD at this location is typically deepest in the winters and shallowest towards the end of the spring/summer upwelling season. Samples of 30 m were both within and below the ML, and the site shows high MLD variability.
Fig. S8. Clustering of hybridizations by sample and by genotype, per Fig. 4, using only the subset of the 30 m samples definitively below the mixed layer depth (MLD). MLD is shown in Fig. S7 and was calculated as the first depth (≥ 10 m) with > 0.1°C difference from the previous meter (per MBARI BOG group, R. Michisaki, pers. comm.). Excluding the 30 m samples above the MLD does not result in discrete clustering of the 0 m and 30 m samples.
Fig. S9. Array profiles for all targets within three common phylogenetic clades: (A) *Roseobacter*, (B) SAR86, (C) SAR11.
Fig. S10. Heatmap of array hybridizations with samples ordered chronologically, without clustering of samples (columns) or genotypes (rows). The break between the 2000–2002 and 2003–2004 sampling periods is indicated by the black vertical dashed line. Intensity of cell colour indicates relative target signal for that genotype and sample date; note that relative abundance is quantitative for each genotype between samples but not between genotypes. Samples are named Depth_Year_CollectionDate, and are colour-coded by oceanographic season (see colour legend and text). Red asterisks denote samples with particularly intense 0 m profiles. Grey columns indicate no samples for that depth and date. (A) 0 m samples, (B) 30 m samples, (C) 200 m samples, with the three depths vertically stacked.
Fig. S11. Evaluating the genetic relatedness of community DNA hybridized to the array. On the left are mean organism signals as shown in Fig. 4, repeated here for side-by-side examination. On the right are the relative ratios of the Tukey biweights (TBW) to the means for each organism (samples in same order as clustering based on mean signals, on left). This ratio is related to the identity of hybridized DNA to the target sequence. Hybridized DNAs with a large relative drop in signal when assessed as TBW rather than as mean (darker blue) have a less even signal across their target probe sets, and are thus inferred to be less closely related to the target sequence (i.e. 80–90% ANI), whereas hybridized DNAs with higher TBW:Mean ratios (lighter blue) are inferred to be genotypes more closely related to targeted sequences (i.e. > 90% ANI), as in Rich and colleagues (2008).
Table S1. Array targets
Table S2. Array targets summarized by phylogenetic cluster
Table S3. Comparison of array with other broad taxonomic surveys of Monterey Bay.
Table S4. Nutrient data for the sample site (Station M1) 2000–2004.

Please note: Wiley-Blackwell are not responsible for the content or functionality of any supporting materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.

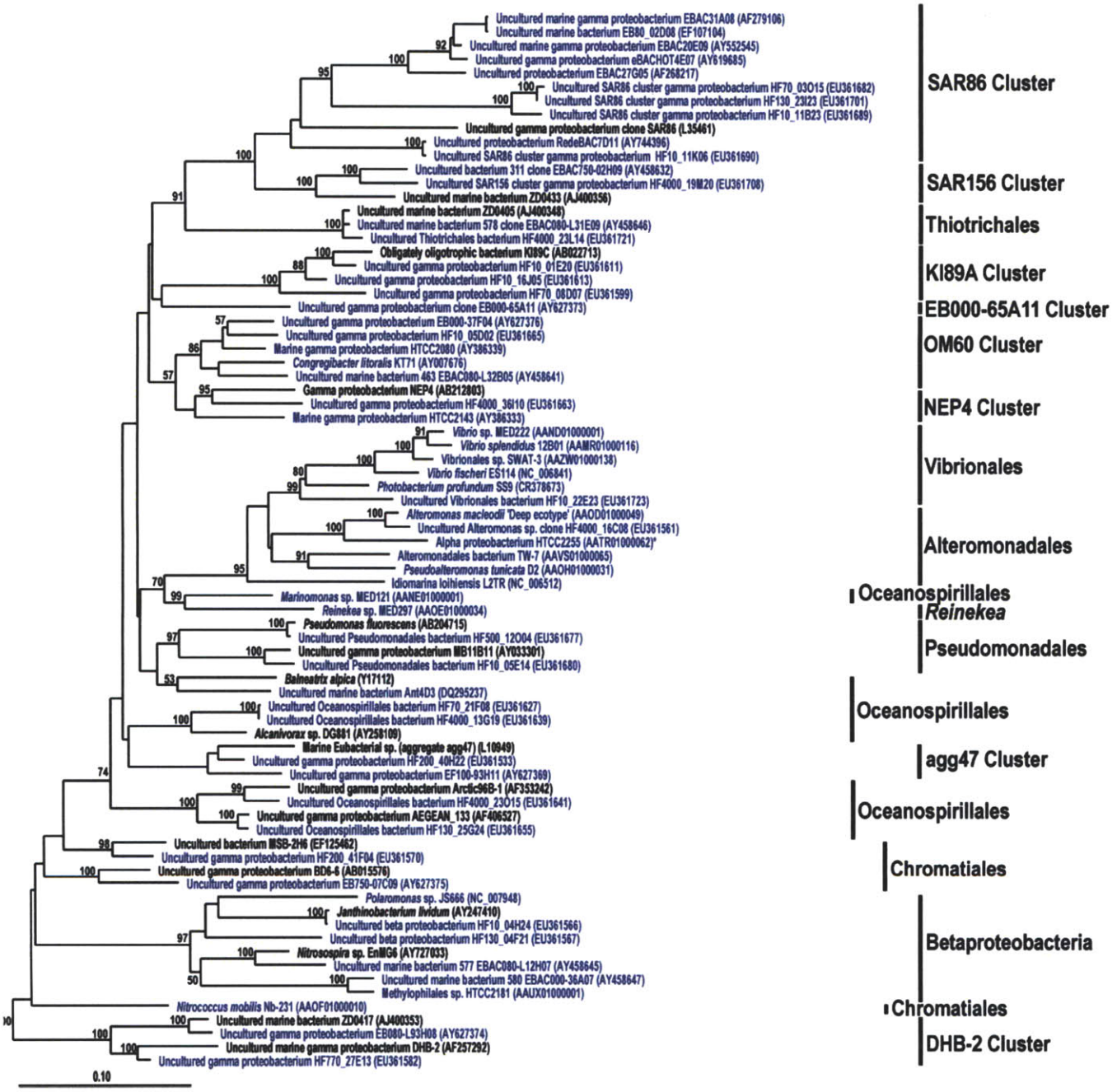

Uncultured marine gamma proteobacterium EBAC31A08 (AF279106)
Uncultured marine bacterium EB80_02D08 (EF107104)
Uncultured marine gamma proteobacterium EBAC20E09 (AY552545)
Uncultured gamma proteobacterium eBACHOT4E07 (AY619685)
Uncultured proteobacterium EBAC27G05 (AF268217)
Uncultured SAR86 cluster gamma proteobacterium HF70_03O15 (EU361682)
Uncultured SAR86 cluster gamma proteobacterium HF130_23I23 (EU361701)
Uncultured SAR86 cluster gamma proteobacterium HF10_11B23 (EU361689)
Uncultured gamma proteobacterium clone SAR86 (L35461)
Uncultured proteobacterium RedeBAC7D11 (AY744396)
Uncultured SAR86 cluster gamma proteobacterium HF10_11K06 (EU361690)
Uncultured bacterium 311 clone EBAC750-02H09 (AY458632)
Uncultured SAR156 cluster gamma proteobacterium HF4000_19M20 (EU361708)
Uncultured marine bacterium ZD0433 (AJ400356)
Uncultured marine bacterium ZD0405 (AJ400348)
Uncultured marine bacterium 578 clone EBAC080-L31E09 (AY458646)
Uncultured Thiotrichales bacterium HF4000_23L14 (EU361721)
Obligately oligotrophic bacterium KI89C (AB022713)
Uncultured gamma proteobacterium HF10_01E20 (EU361611)
Uncultured gamma proteobacterium HF10_16J05 (EU361613)
Uncultured gamma proteobacterium HF70_08D07 (EU361599)
Uncultured gamma proteobacterium clone EB000-65A11 (AY627373)
Uncultured gamma proteobacterium EB000-37F04 (AY627376)
Uncultured gamma proteobacterium HF10_05D02 (EU361665)
Marine gamma proteobacterium HTCC2080 (AY386339)
Congregibacter litoralis KT71 (AY007676)
Uncultured marine bacterium 463 EBAC080-L32B05 (AY458641)
Gamma proteobacterium NEP4 (AB212803)
Uncultured gamma proteobacterium HF4000_36I10 (EU361663)
Marine gamma proteobacterium HTCC2143 (AY386333)
Vibrio sp. MED222 (AAND01000001)
Vibrio splendidus 12B01 (AAMR01000116)
Vibrionales sp. SWAT-3 (AAZW01000138)
Vibrio fischeri ES114 (NC_006841)
Photobacterium profundum SS9 (CR378673)
Uncultured Vibrionales bacterium HF10_22E23 (EU361723)
Alteromonas macleodii 'Deep ecotype' (AAOD01000049)
Uncultured Alteromonas sp. clone HF4000_16C08 (EU361561)
Alpha proteobacterium HTCC2255 (AATR01000062)*
Alteromonadales bacterium TW-7 (AAVS01000065)
Pseudoalteromonas tunicata D2 (AAOH01000031)
Idiomarina loihiensis L2TR (NC_006512)
Marinomonas sp. MED121 (AANE01000001)
Reinekea sp. MED297 (AAOE01000034)
Pseudomonas fluorescens (AB204715)
Uncultured Pseudomonadales bacterium HF500_12O04 (EU361677)
Uncultured gamma proteobacterium MB11B11 (AY033301)
Uncultured Pseudomonadales bacterium HF10_05E14 (EU361680)
Balneatrix alpica (Y17112)
Uncultured marine bacterium Ant4D3 (DQ295237)
Uncultured Oceanospirillales bacterium HF70_21F08 (EU361627)
Uncultured Oceanospirillales bacterium HF4000_13G19 (EU361639)
Alcanivorax sp. DG881 (AY258109)
Marine Eubacterial sp. (aggregate agg47) (L10949)
Uncultured gamma proteobacterium HF200_40H22 (EU361533)
Uncultured gamma proteobacterium EF100-93H11 (AY627369)
Uncultured gamma proteobacterium Arctic96B-1 (AF353242)
Uncultured Oceanospirillales bacterium HF4000_23O15 (EU361641)
Uncultured gamma proteobacterium AEGEAN_133 (AF406527)
Uncultured Oceanospirillales bacterium HF130_25G24 (EU361655)
Uncultured bacterium MSB-2H6 (EF125462)
Uncultured gamma proteobacterium HF200_41F04 (EU361570)
Uncultured gamma proteobacterium BD6-6 (AB015576)
Uncultured gamma proteobacterium EB750-07C09 (AY627375)
Polaromonas sp. JS666 (NC_007948)
Janthinobacterium lividum (AY247410)
Uncultured beta proteobacterium HF10_04H24 (EU361566)
Uncultured beta proteobacterium HF130_04F21 (EU361567)
Nitrosospira sp. EnMG6 (AY727033)
Uncultured marine bacterium 577 EBAC080-L12H07 (AY458645)
Uncultured marine bacterium 580 EBAC000-36A07 (AY458647)
Methylophilales sp. HTCC2181 (AAUX01000001)
Nitrococcus mobilis Nb-231 (AAOF01000010)
Uncultured marine bacterium ZD0417 (AJ400353)
Uncultured gamma proteobacterium EB080-L93H08 (AY627374)
Uncultured marine gamma proteobacterium DHB-2 (AF257292)
Uncultured gamma proteobacterium HF770_27E13 (EU361582)
0.10
SAR86 Cluster
SAR156 Cluster
Thiotrichales
KI89A Cluster
EB000-65A11 Cluster
OM60 Cluster
NEP4 Cluster
Vibrionales
Alteromonadales
Oceanospirillales
Reinekea
Pseudomonadales
Oceanospirillales
agg47 Cluster
Oceanospirillales
Chromatiales
Betaproteobacteria
Chromatiales
DHB-2 Cluster

Figure S1

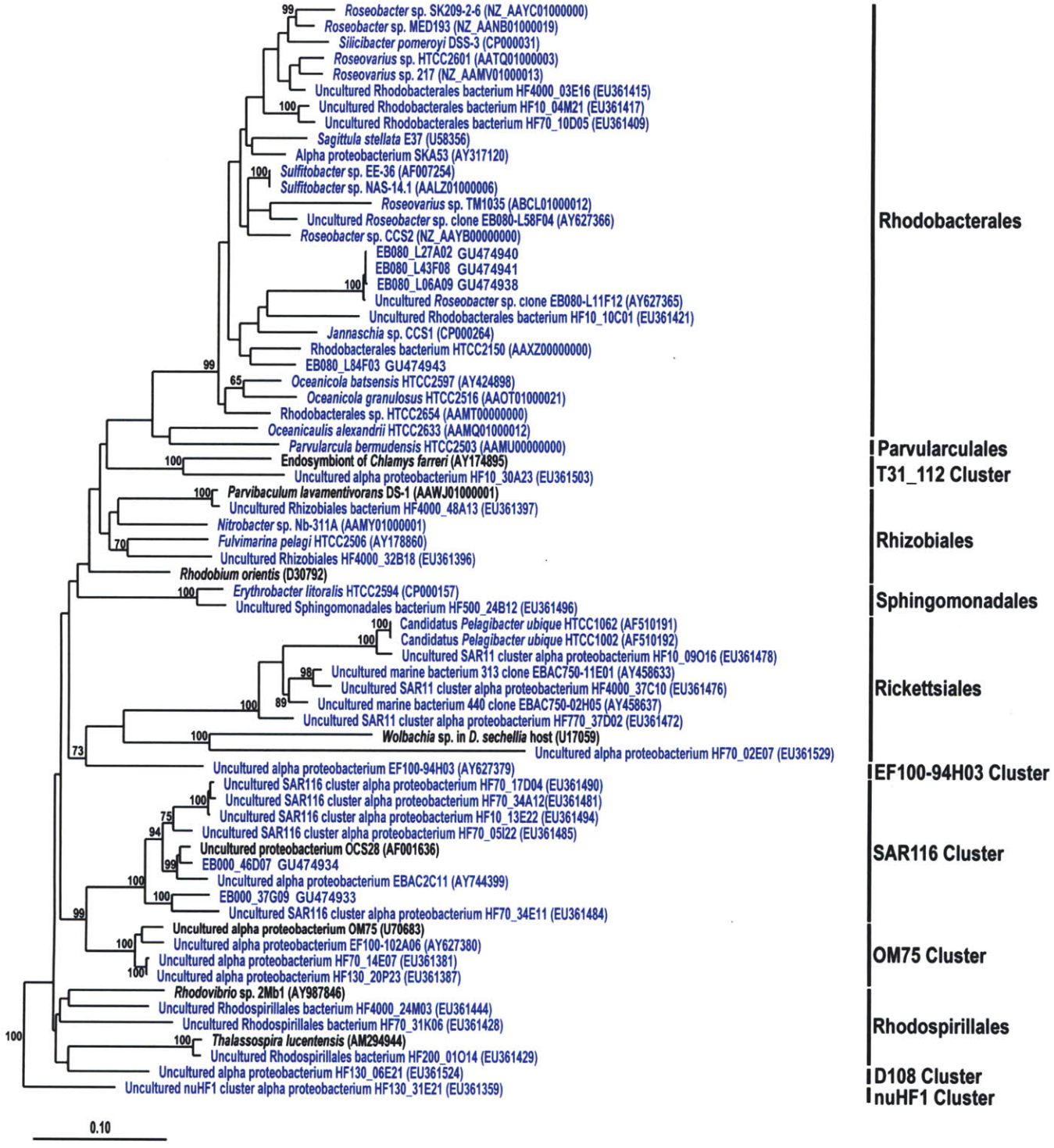

Roseobacter sp. SK209-2-6 (NZ_AAYC01000000)
Roseobacter sp. MED193 (NZ_AANB01000019)
Silicibacter pomeroyi DSS-3 (CP000031)
Roseovarius sp. HTCC2601 (AATQ01000003)
Roseovarius sp. 217 (NZ_AAMV01000013)
Uncultured Rhodobacterales bacterium HF4000_03E16 (EU361415)
Uncultured Rhodobacterales bacterium HF10_04M21 (EU361417)
Uncultured Rhodobacterales bacterium HF70_10D05 (EU361409)
Sagittula stellata E37 (U58356)
Alpha proteobacterium SKA53 (AY317120)
Sulfitobacter sp. EE-36 (AF007254)
Sulfitobacter sp. NAS-14.1 (AALZ01000006)
Roseovarius sp. TM1035 (ABCL01000012)
Uncultured Roseobacter sp. clone EB080-L58F04 (AY627366)
Roseobacter sp. CCS2 (NZ_AAYB00000000)
EB080_L27A02 GU474940
EB080_L43F08 GU474941
EB080_L06A09 GU474938
Uncultured Roseobacter sp. clone EB080-L11F12 (AY627365)
Uncultured Rhodobacterales bacterium HF10_10C01 (EU361421)
Jannaschia sp. CCS1 (CP000264)
Rhodobacterales bacterium HTCC2150 (AAXZ00000000)
EB080_L84F03 GU474943
Oceanicola batsensis HTCC2597 (AY424898)
Oceanicola granulosus HTCC2516 (AAOT01000021)
Rhodobacterales sp. HTCC2654 (AAMT00000000)
Oceanicaulis alexandrii HTCC2633 (AAMQ01000012)
Parvularcula bermudensis HTCC2503 (AAMU00000000)
Endosymbiont of Chlamys farreri (AY174895)
Uncultured alpha proteobacterium HF10_30A23 (EU361503)
Parvibaculum lavamentivorans DS-1 (AAWJ01000001)
Uncultured Rhizobiales bacterium HF4000_48A13 (EU361397)
Nitrobacter sp. Nb-311A (AAMY01000001)
Fulvimarina pelagi HTCC2506 (AY178860)
Uncultured Rhizobiales HF4000_32B18 (EU361396)
Rhodobium orientis (D30792)
Erythrobacter litoralis HTCC2594 (CP000157)
Uncultured Sphingomonadales bacterium HF500_24B12 (EU361496)
Candidatus Pelagibacter ubique HTCC1062 (AF510191)
Candidatus Pelagibacter ubique HTCC1002 (AF510192)
Uncultured SAR11 cluster alpha proteobacterium HF10_09O16 (EU361478)
Uncultured marine bacterium 313 clone EBAC750-11E01 (AY458633)
Uncultured SAR11 cluster alpha proteobacterium HF4000_37C10 (EU361476)
Uncultured marine bacterium 440 clone EBAC750-02H05 (AY458637)
Uncultured SAR11 cluster alpha proteobacterium HF770_37D02 (EU361472)
Wolbachia sp. in D. sechellia host (U17059)
Uncultured alpha proteobacterium HF70_02E07 (EU361529)
Uncultured alpha proteobacterium EF100-94H03 (AY627379)
Uncultured SAR116 cluster alpha proteobacterium HF70_17D04 (EU361490)
Uncultured SAR116 cluster alpha proteobacterium HF70_34A12(EU361481)
Uncultured SAR116 cluster alpha proteobacterium HF10_13E22 (EU361494)
Uncultured SAR116 cluster alpha proteobacterium HF70_05I22 (EU361485)
Uncultured proteobacterium OCS28 (AF001636)
EB000_46D07 GU474934
Uncultured alpha proteobacterium EBAC2C11 (AY744399)
EB000_37G09 GU474933
Uncultured SAR116 cluster alpha proteobacterium HF70_34E11 (EU361484)
Uncultured alpha proteobacterium OM75 (U70683)
Uncultured alpha proteobacterium EF100-102A06 (AY627380)
Uncultured alpha proteobacterium HF70_14E07 (EU361381)
Uncultured alpha proteobacterium HF130_20P23 (EU361387)
Rhodovibrio sp. 2Mb1 (AY987846)
Uncultured Rhodospirillales bacterium HF4000_24M03 (EU361444)
Uncultured Rhodospirillales bacterium HF70_31K06 (EU361428)
Thalassospira lucentensis (AM294944)
Uncultured Rhodospirillales bacterium HF200_01O14 (EU361429)
Uncultured alpha proteobacterium HF130_06E21 (EU361524)
Uncultured nuHF1 cluster alpha proteobacterium HF130_31E21 (EU361359)
0.10
Rhodobacterales
Parvularculales
T31_112 Cluster
Rhizobiales
Sphingomonadales
Rickettsiales
EF100-94H03 Cluster
SAR116 Cluster
OM75 Cluster
Rhodospirillales
D108 Cluster
nuHF1 Cluster

Figure S2

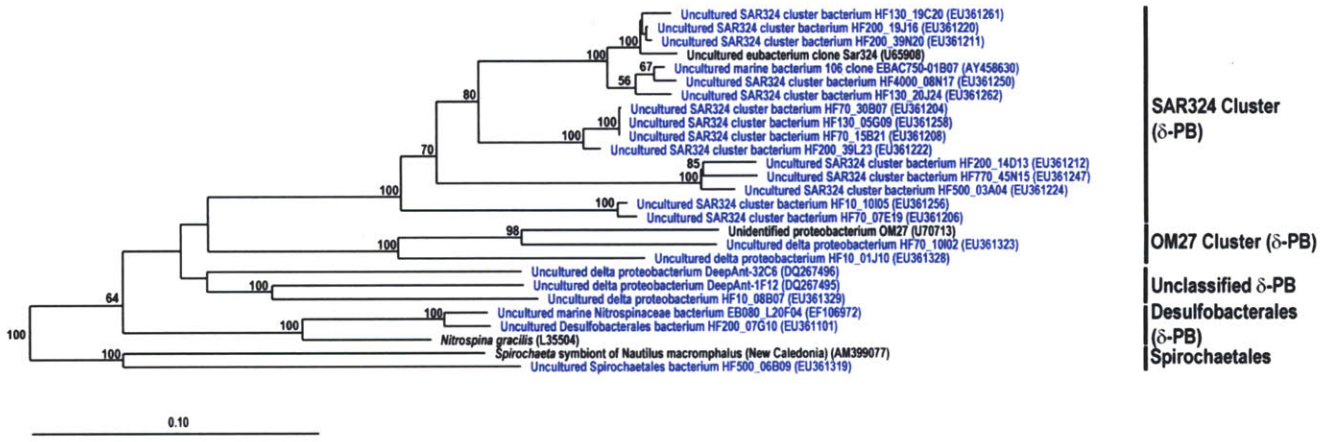

Uncultured SAR324 cluster bacterium HF130_19C20 (EU361261)
Uncultured SAR324 cluster bacterium HF200_19J16 (EU361220)
Uncultured SAR324 cluster bacterium HF200_39N20 (EU361211)
Uncultured eubacterium clone Sar324 (U65906)
Uncultured marine bacterium 106 clone EBAC750-01B07 (AY458630)
Uncultured SAR324 cluster bacterium HF4000_08N17 (EU361250)
Uncultured SAR324 cluster bacterium HF130_20J24 (EU361262)
Uncultured SAR324 cluster bacterium HF70_30B07 (EU361204)
Uncultured SAR324 cluster bacterium HF130_05G09 (EU361258)
Uncultured SAR324 cluster bacterium HF70_15B21 (EU361208)
Uncultured SAR324 cluster bacterium HF200_39L23 (EU361222)
Uncultured SAR324 cluster bacterium HF200_14D13 (EU361212)
Uncultured SAR324 cluster bacterium HF770_45N15 (EU361247)
Uncultured SAR324 cluster bacterium HF500_03A04 (EU361224)
Uncultured SAR324 cluster bacterium HF10_10I05 (EU361256)
Uncultured SAR324 cluster bacterium HF70_07E19 (EU361206)
Unidentified proteobacterium OM27 (U70713)
Uncultured delta proteobacterium HF70_10I02 (EU361323)
Uncultured delta proteobacterium HF10_01J10 (EU361328)
Uncultured delta proteobacterium DeepAnt-32C6 (DQ267496)
Uncultured delta proteobacterium DeepAnt-1F12 (DQ267495)
Uncultured delta proteobacterium HF10_08B07 (EU361329)
Uncultured marine Nitrospinaceae bacterium EB080_L20F04 (EF106972)
Uncultured Desulfobacterales bacterium HF200_07G10 (EU361101)
Nitrospina gracilis (L35504)
Spirochaeta symbiont of Nautilus macromphalus (New Caledonia) (AM399077)
Uncultured Spirochaetales bacterium HF500_06B09 (EU361319)
0.10
SAR324 Cluster (δ-PB)
OM27 Cluster (δ-PB)
Unclassified δ-PB
Desulfobacterales (δ-PB)
Spirochaetales

Figure S3

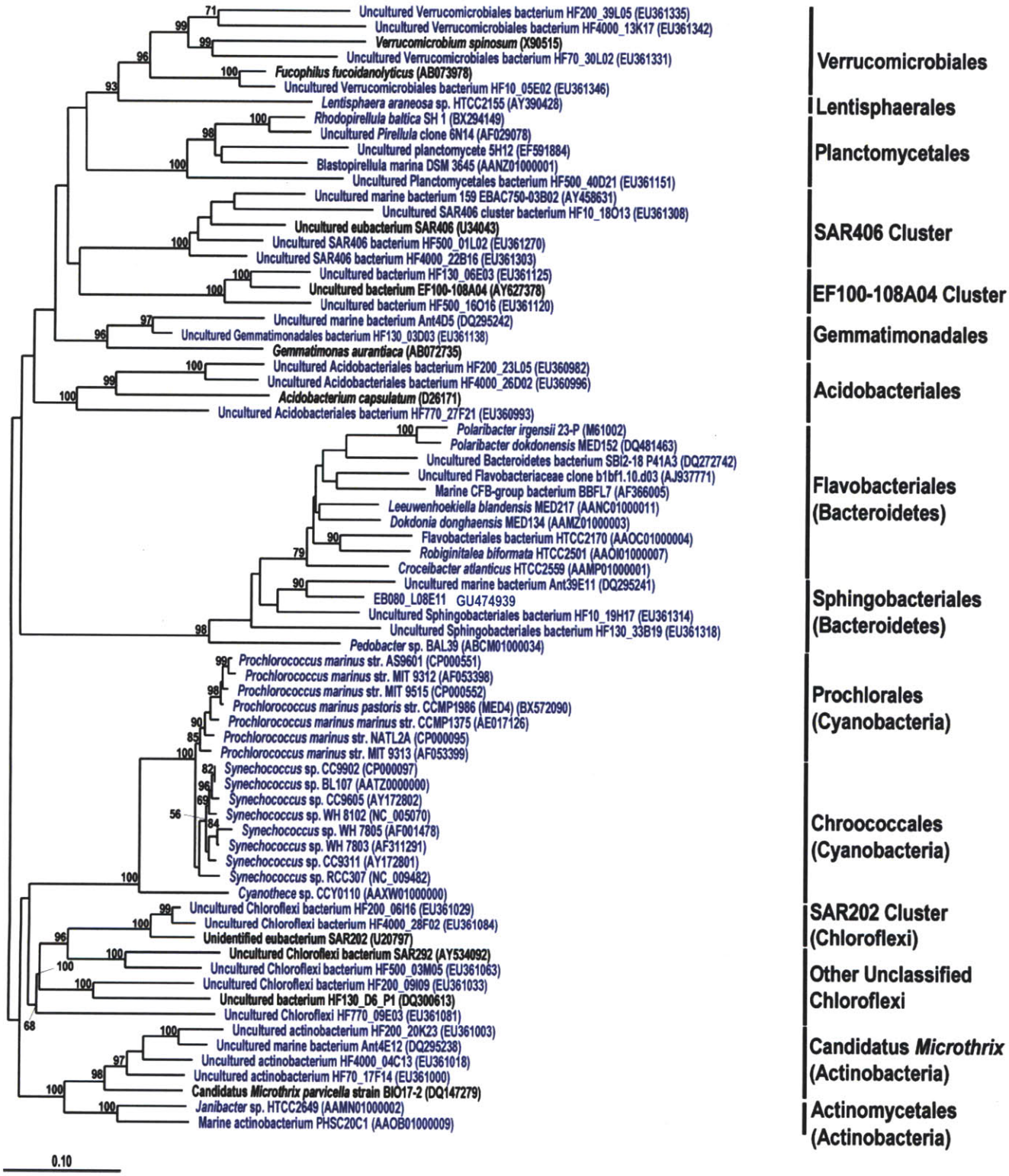

Uncultured Verrucomicrobiales bacterium HF200_39L05 (EU361335)
Uncultured Verrucomicrobiales bacterium HF4000_13K17 (EU361342)
Verrucomicrobium spinosum (X90515)
Uncultured Verrucomicrobiales bacterium HF70_30L02 (EU361331)
Fucophilus fucoidanolyticus (AB073978)
Uncultured Verrucomicrobiales bacterium HF10_05E02 (EU361346)
Lentisphaera araneosa sp. HTCC2155 (AY390428)
Rhodopirellula baltica SH 1 (BX294149)
Uncultured Pirellula clone 6N14 (AF029078)
Uncultured planctomycete 5H12 (EF591884)
Blastopirellula marina DSM 3645 (AANZ01000001)
Uncultured Planctomycetales bacterium HF500_40D21 (EU361151)
Uncultured marine bacterium 159 EBAC750-03B02 (AY458631)
Uncultured SAR406 cluster bacterium HF10_18O13 (EU361308)
Uncultured eubacterium SAR406 (U34043)
Uncultured SAR406 bacterium HF500_01L02 (EU361270)
Uncultured SAR406 bacterium HF4000_22B16 (EU361303)
Uncultured bacterium HF130_06E03 (EU361125)
Uncultured bacterium EF100-108A04 (AY627378)
Uncultured bacterium HF500_16O16 (EU361120)
Uncultured marine bacterium Ant4D5 (DQ295242)
Uncultured Gemmatimonadales bacterium HF130_03D03 (EU361138)
Gemmatimonas aurantiaca (AB072735)
Uncultured Acidobacteriales bacterium HF200_23L05 (EU360982)
Uncultured Acidobacteriales bacterium HF4000_26D02 (EU360996)
Acidobacterium capsulatum (D26171)
Uncultured Acidobacteriales bacterium HF770_27F21 (EU360993)
Polaribacter irgensii 23-P (M61002)
Polaribacter dokdonensis MED152 (DQ481463)
Uncultured Bacteroidetes bacterium SBI2-18 P41A3 (DQ272742)
Uncultured Flavobacteriaceae clone b1bf1.10.d03 (AJ937771)
Marine CFB-group bacterium BBFL7 (AF366005)
Leeuwenhoekiella blandensis MED217 (AANC01000011)
Dokdonia donghaensis MED134 (AAMZ01000003)
Flavobacteriales bacterium HTCC2170 (AAOC01000004)
Robiginitalea biformata HTCC2501 (AAOI01000007)
Croceibacter atlanticus HTCC2559 (AAMP01000001)
Uncultured marine bacterium Ant39E11 (DQ295241)
EB080_L08E11 GU474939
Uncultured Sphingobacteriales bacterium HF10_19H17 (EU361314)
Uncultured Sphingobacteriales bacterium HF130_33B19 (EU361318)
Pedobacter sp. BAL39 (ABCM01000034)
Prochlorococcus marinus str. AS9601 (CP000551)
Prochlorococcus marinus str. MIT 9312 (AF053398)
Prochlorococcus marinus str. MIT 9515 (CP000552)
Prochlorococcus marinus pastoris str. CCMP1986 (MED4) (BX572090)
Prochlorococcus marinus marinus str. CCMP1375 (AE017126)
Prochlorococcus marinus str. NATL2A (CP000095)
Prochlorococcus marinus str. MIT 9313 (AF053399)
Synechococcus sp. CC9902 (CP000097)
Synechococcus sp. BL107 (AATZ0000000)
Synechococcus sp. CC9605 (AY172802)
Synechococcus sp. WH 8102 (NC_005070)
Synechococcus sp. WH 7805 (AF001478)
Synechococcus sp. WH 7803 (AF311291)
Synechococcus sp. CC9311 (AY172801)
Synechococcus sp. RCC307 (NC_009482)
Cyanothece sp. CCY0110 (AAXW01000000)
Uncultured Chloroflexi bacterium HF200_06I16 (EU361029)
Uncultured Chloroflexi bacterium HF4000_28F02 (EU361084)
Unidentified eubacterium SAR202 (U20797)
Uncultured Chloroflexi bacterium SAR292 (AY534092)
Uncultured Chloroflexi bacterium HF500_03M05 (EU361063)
Uncultured Chloroflexi bacterium HF200_09I09 (EU361033)
Uncultured bacterium HF130_D6_P1 (DQ300613)
Uncultured Chloroflexi HF770_09E03 (EU361081)
Uncultured actinobacterium HF200_20K23 (EU361003)
Uncultured marine bacterium Ant4E12 (DQ295238)
Uncultured actinobacterium HF4000_04C13 (EU361018)
Uncultured actinobacterium HF70_17F14 (EU361000)
Candidatus Microthrix parvicella strain BIO17-2 (DQ147279)
Janibacter sp. HTCC2649 (AAMN01000002)
Marine actinobacterium PHSC20C1 (AAOB01000009)
0.10
Verrucomicrobiales
Lentisphaerales
Planctomycetales
SAR406 Cluster
EF100-108A04 Cluster
Gemmatimonadales
Acidobacteriales
Flavobacteriales (Bacteroidetes)
Sphingobacteriales (Bacteroidetes)
Prochlorales (Cyanobacteria)
Chroococcales (Cyanobacteria)
SAR202 Cluster (Chloroflexi)
Other Unclassified Chloroflexi
Candidatus Microthrix (Actinobacteria)
Actinomycetales (Actinobacteria)

Figure S4


Uncultured marine group II euryarchaeote DeepAnt-JyKC7 clone (AY53410)
Uncultured marine group II euryarchaeote EF100_57A08 (EU221238)
Uncultured marine group II euryarchaeote HF70_59C08 (DQ156348)
Uncultured euryarchaeote Alv-FOS1 (DQ118403)
Uncultured euryarchaeote Alv-FOS4 (DQ118404)
Uncultured euryarchaeote Alv-FOS5 (DQ078753)
Thermococcus barophilus DSM 11836 (AY099172)
Uncultured crenarchaeote DeepAnt-EC39 (AY316120)
Uncultured marine group I crenarchaeote HF500_25H11 (DQ300517)
Uncultured crenarchaeote 4B7 (U40238)
Uncultured crenarchaeote 31B02 (AF393306)
Uncultured crenarchaeote 74A4 (AF393466)
100
100
80
100
88
97
89
100
100
Marine Uncultured Group II (Euryarchaeota)
Unclassified Euryarchaeota
Marine Uncultured Group I (Crenarchaeota)
0.10

Figure S5

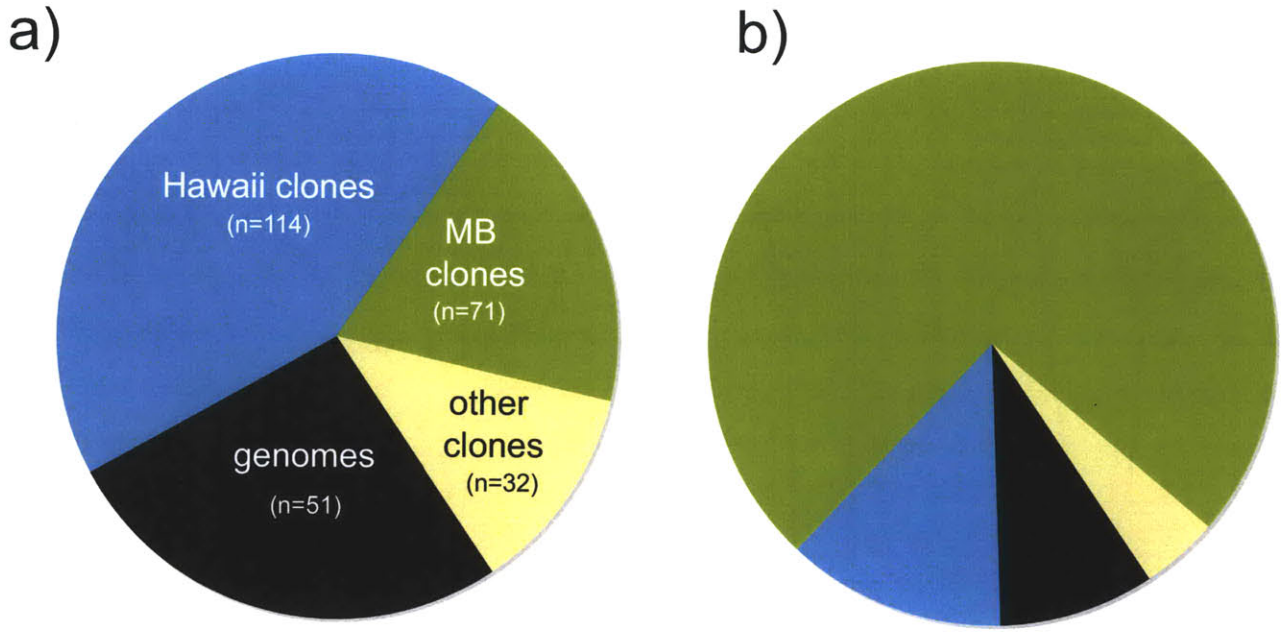

a)
Hawaii clones
(n=114)
MB
clones
(n=71)
other
clones
(n=32)
genomes
(n=51)
b)

Figure S6

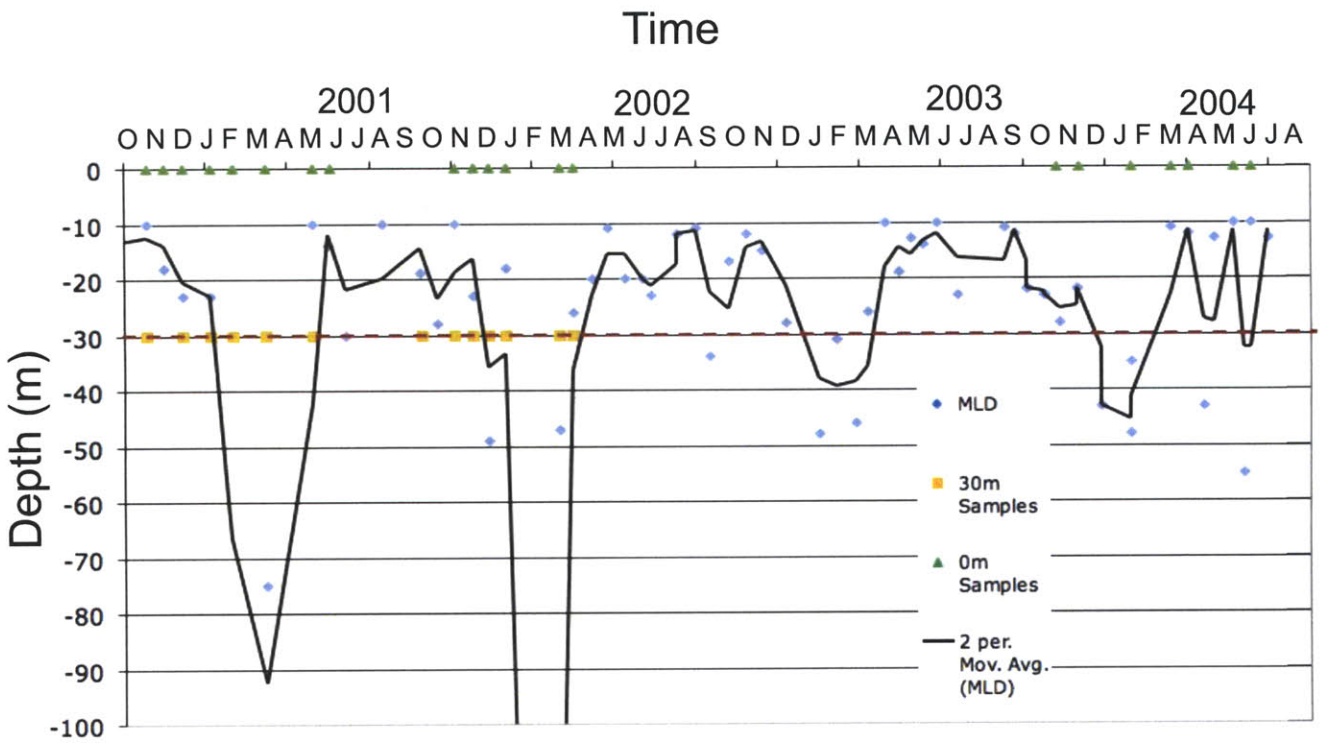

Time
2001
2002
2003
2004
O N D J F M A M J J A S O N D J F M A M J J A S O N D J F M A M J J A S O N D J F M A M J J A
0
-10
-20
-30
-40
-50
-60
-70
-80
-90
-100
Depth (m)
MLD
30m Samples
0m Samples
2 per. Mov. Avg. (MLD)

Figure S7

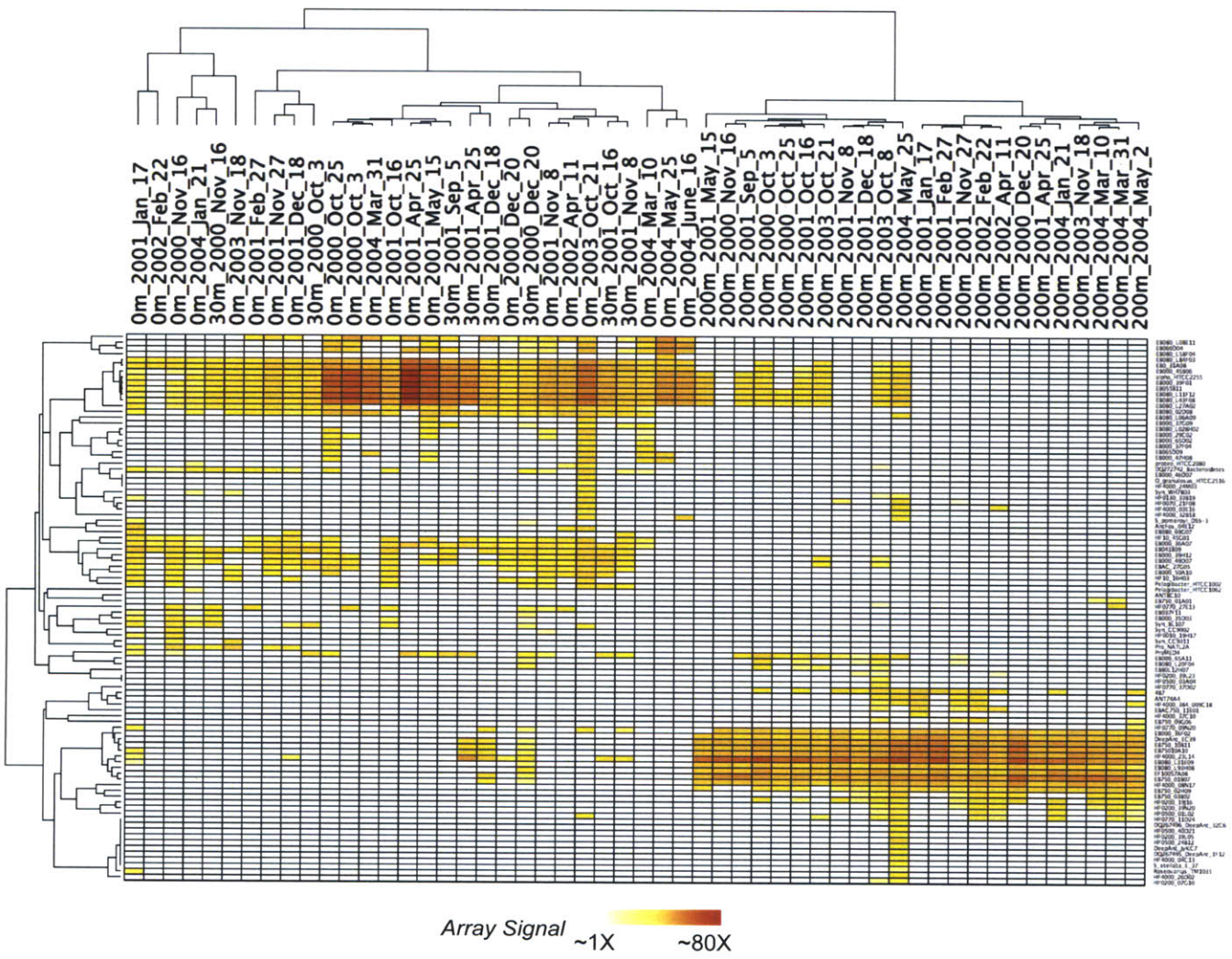

0m_2001_Jan_17
0m_2002_Feb_22
0m_2000_Nov_16
0m_2004_Jan_21
30m_2000_Nov_16
0m_2003_Nov_18
0m_2001_Feb_27
0m_2001_Nov_27
0m_2001_Dec_18
30m_2000_Oct_3
0m_2000_Oct_25
0m_2000_Oct_3
0m_2004_Mar_31
0m_2001_Oct_16
0m_2001_Apr_25
0m_2001_May_15
30m_2001_Sep_5
30m_2001_Apr_25
30m_2001_Dec_18
0m_2000_Dec_20
30m_2000_Dec_20
0m_2001_Nov_8
0m_2002_Apr_11
0m_2003_Oct_21
30m_2001_Oct_16
30m_2001_Nov_8
0m_2004_Mar_10
0m_2004_May_25
0m_2004_June_16
200m_2001_May_15
200m_2000_Nov_16
200m_2001_Sep_5
200m_2000_Oct_3
200m_2000_Oct_25
200m_2001_Oct_16
200m_2003_Oct_21
200m_2001_Nov_8
200m_2001_Dec_18
200m_2003_Oct_8
200m_2004_May_25
200m_2001_Jan_17
200m_2001_Feb_27
200m_2001_Nov_27
200m_2002_Feb_22
200m_2002_Apr_11
200m_2000_Dec_20
200m_2001_Apr_25
200m_2004_Jan_21
200m_2003_Nov_18
200m_2004_Mar_10
200m_2004_Mar_31
200m_2004_May_2
Array Signal
~1X
~80X

Figure S8

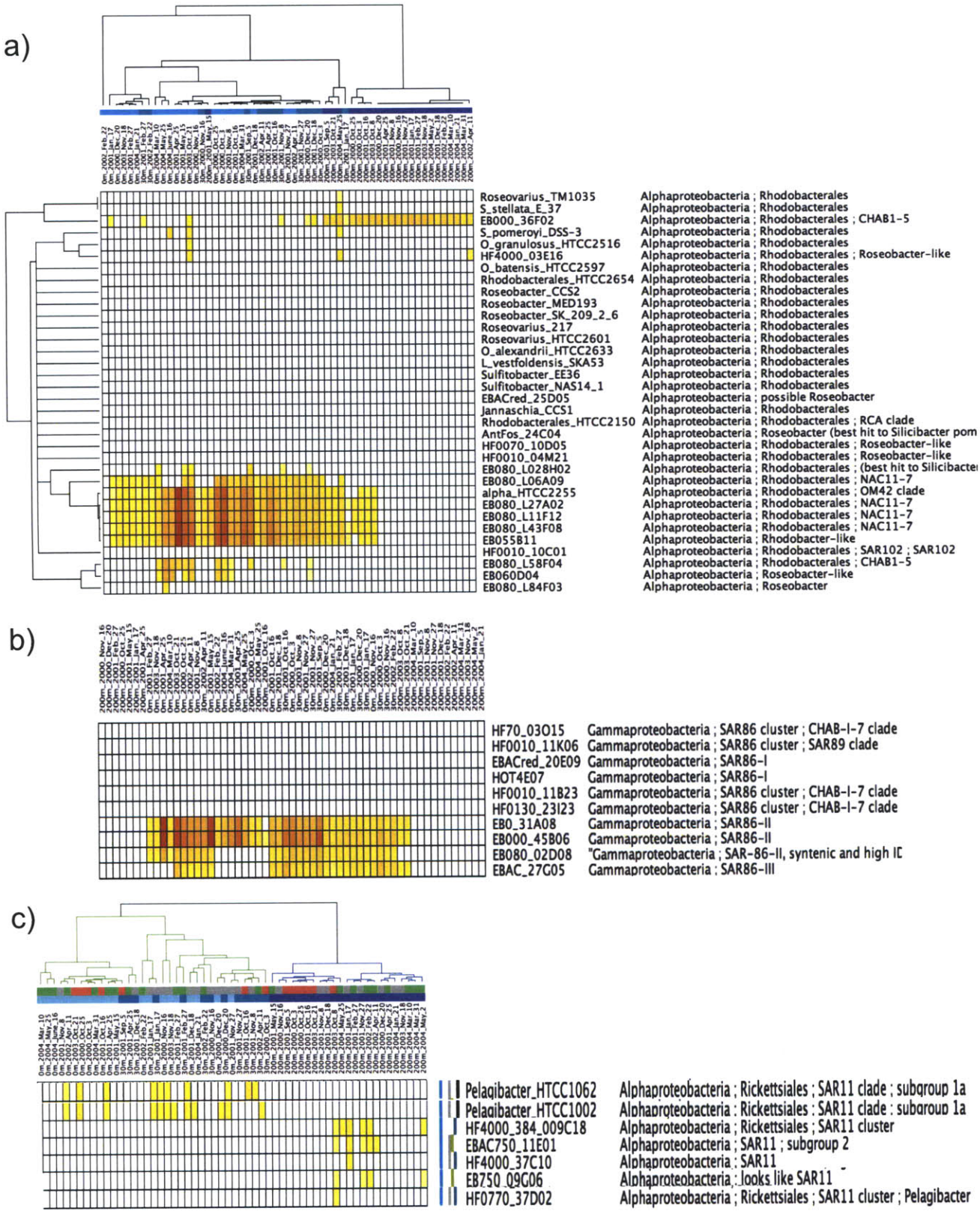

a)
Roseovarius_TM1035 Alphaproteobacteria ; Rhodobacterales
S_stellata_E_37 Alphaproteobacteria ; Rhodobacterales
EB000_36F02 Alphaproteobacteria ; Rhodobacterales ; CHAB1-5
S_pomeroyi_DSS-3 Alphaproteobacteria ; Rhodobacterales
O_granulosus_HTCC2516 Alphaproteobacteria ; Rhodobacterales
HF4000_03E16 Alphaproteobacteria ; Rhodobacterales ; Roseobacter-like
O_batensis_HTCC2597 Alphaproteobacteria ; Rhodobacterales
Rhodobacterales_HTCC2654 Alphaproteobacteria ; Rhodobacterales
Roseobacter_CCS2 Alphaproteobacteria ; Rhodobacterales
Roseobacter_MED193 Alphaproteobacteria ; Rhodobacterales
Roseobacter_SK_209_2_6 Alphaproteobacteria ; Rhodobacterales
Roseovarius_217 Alphaproteobacteria ; Rhodobacterales
Roseovarius_HTCC2601 Alphaproteobacteria ; Rhodobacterales
O_alexandrii_HTCC2633 Alphaproteobacteria ; Rhodobacterales
L_vestfoldensis_SKA53 Alphaproteobacteria ; Rhodobacterales
Sulfitobacter_EE36 Alphaproteobacteria ; Rhodobacterales
Sulfitobacter_NAS14_1 Alphaproteobacteria ; Rhodobacterales
EBACred_25D05 Alphaproteobacteria ; possible Roseobacter
Jannaschia_CCS1 Alphaproteobacteria ; Rhodobacterales
Rhodobacterales_HTCC2150 Alphaproteobacteria ; Rhodobacterales ; RCA clade
AntFos_24C04 Alphaproteobacteria ; Roseobacter (best hit to Silicibacter pom
HF0070_10D05 Alphaproteobacteria ; Rhodobacterales ; Roseobacter-like
HF0010_04M21 Alphaproteobacteria ; Rhodobacterales ; Roseobacter-like
EB080_L028H02 Alphaproteobacteria ; Rhodobacterales ; (best hit to Silicibacter
EB080_L06A09 Alphaproteobacteria ; Rhodobacterales ; NAC11-7
alpha_HTCC2255 Alphaproteobacteria ; Rhodobacterales ; OM42 clade
EB080_L27A02 Alphaproteobacteria ; Rhodobacterales ; NAC11-7
EB080_L11F12 Alphaproteobacteria ; Rhodobacterales ; NAC11-7
EB080_L43F08 Alphaproteobacteria ; Rhodobacterales ; NAC11-7
EB055B11 Alphaproteobacteria ; Rhodobacter-like
HF0010_10C01 Alphaproteobacteria ; Rhodobacterales ; SAR102 ; SAR102
EB080_L58F04 Alphaproteobacteria ; Rhodobacterales ; CHAB1-5
EB060D04 Alphaproteobacteria ; Roseobacter-like
EB080_L84F03 Alphaproteobacteria ; Roseobacter
b)
HF70_03O15 Gammaproteobacteria ; SAR86 cluster ; CHAB-I-7 clade
HF0010_11K06 Gammaproteobacteria ; SAR86 cluster ; SAR89 clade
EBACred_20E09 Gammaproteobacteria ; SAR86-I
HOT4E07 Gammaproteobacteria ; SAR86-I
HF0010_11B23 Gammaproteobacteria ; SAR86 cluster ; CHAB-I-7 clade
HF0130_23I23 Gammaproteobacteria ; SAR86 cluster ; CHAB-I-7 clade
EB0_31A08 Gammaproteobacteria ; SAR86-II
EB000_45B06 Gammaproteobacteria ; SAR86-II
EB080_02D08 "Gammaproteobacteria ; SAR-86-II, syntenic and high IC
EBAC_27G05 Gammaproteobacteria ; SAR86-III
c)
Pelagibacter_HTCC1062 Alphaproteobacteria ; Rickettsiales ; SAR11 clade ; subgroup 1a
Pelagibacter_HTCC1002 Alphaproteobacteria ; Rickettsiales ; SAR11 clade ; subgroup 1a
HF4000_384_009C18 Alphaproteobacteria ; Rickettsiales ; SAR11 cluster
EBAC750_11E01 Alphaproteobacteria ; SAR11 ; subgroup 2
HF4000_37C10 Alphaproteobacteria ; SAR11
EB750_09G06 Alphaproteobacteria ; looks like SAR11
HF0770_37D02 Alphaproteobacteria ; Rickettsiales ; SAR11 cluster ; Pelagibacter

Figure S9

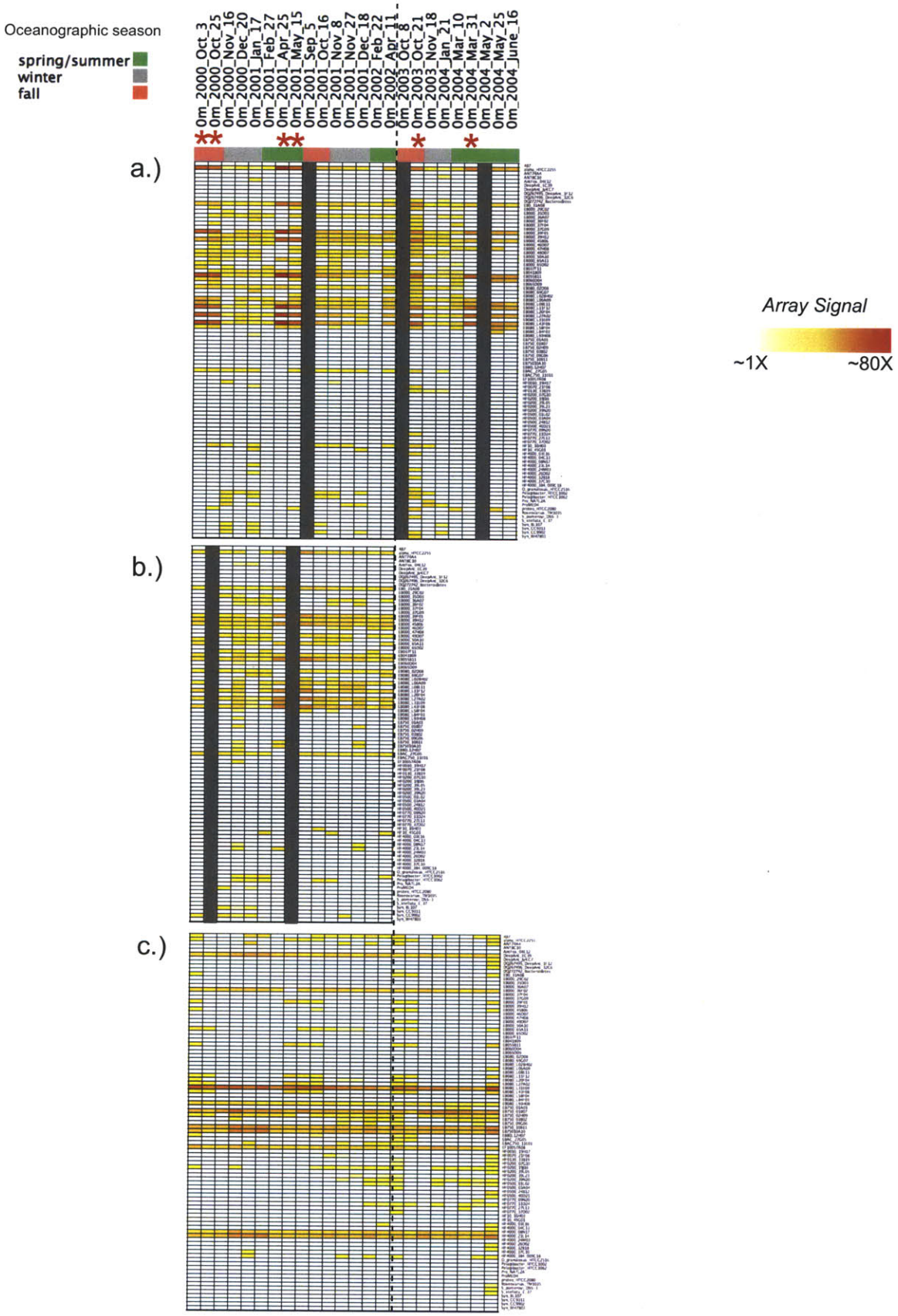

Oceanographic season
spring/summer
winter
fall
0m_2000_Oct_3
0m_2000_Oct_25
0m_2000_Nov_16
0m_2000_Dec_20
0m_2001_Jan_17
0m_2001_Feb_27
0m_2001_Apr_25
0m_2001_May_15
0m_2001_Sep_5
0m_2001_Oct_16
0m_2001_Nov_8
0m_2001_Nov_27
0m_2001_Dec_18
0m_2002_Feb_22
0m_2002_Apr_11
0m_2003_Oct_8
0m_2003_Oct_21
0m_2003_Nov_18
0m_2004_Jan_21
0m_2004_Mar_10
0m_2004_Mar_31
0m_2004_May_2
0m_2004_May_25
0m_2004_June_16
**
**
*
*
a.)
b.)
c.)
Array Signal
~1X
~80X

Figure S10

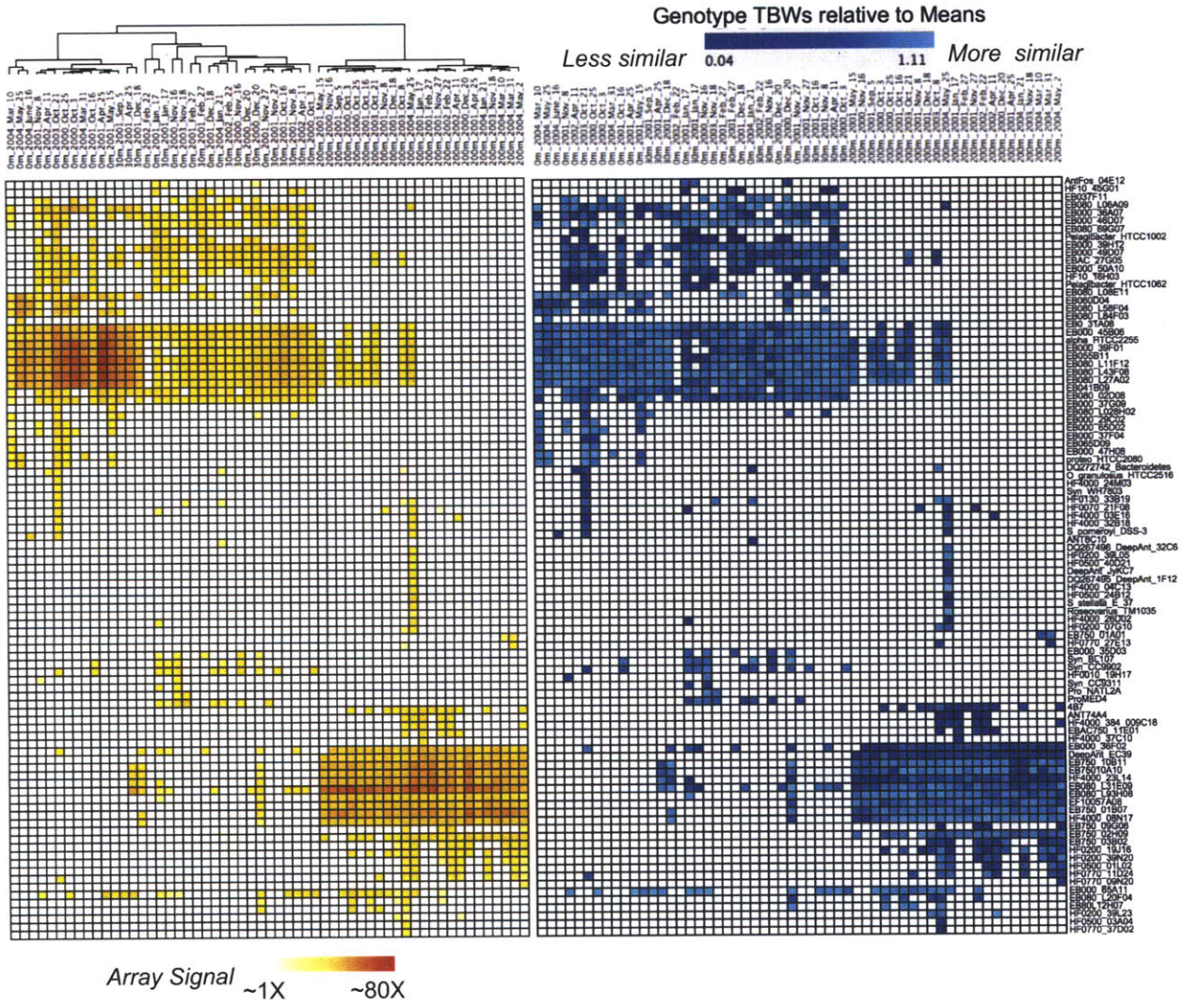

Genotype TBWs relative to Means
Less similar
0.04
1.11
More similar
Array Signal
~1X
~80X

Figure S11

Table S1. Array Targets
(colored to allow easier viewing)

Accession	Target Clone or Genome	Array Probeset Name	Phylogenetic Affilliation
U40238	ORE_4B7	4B7	Archaea ; Crenarchaeota GI
AF393466	ANT_74A4	ANT74A4	Archaea ; Crenarchaeota GI
AF268611	EB000_37F11	EB037F11	Archaea ; Euryarchaeota G2
EU221238	EF100_57A08	EF10057A08	Archaea ; Euryarchaeota G2
DQ257435	HF0010_03D09	HF10_03D09	Archaea ; Euryarchaeota G2
DQ257434	HF0070_19B12	HF70_19B12	Archaea ; Euryarchaeota G2
DQ156348	HF0070_59C08	HF70_59C08	Archaea ; Euryarchaeota G2
EF089401	HF0010_29C11	HF10_29C11	Archaea ; Euryarchaeota G2
AY316120	DeepAnt_EC39	DeepAnt_EC39	Archaea ; Euryarchaeota G2
AY534910	DeepAnt_JyKC7	DeepAnt_JyKC7	Archaea ; Euryarchaeota G2
DQ118403	Alv_FOS1	DQ118403_Alv_FOS1	Archaea ; Euryarchaeota G2
DQ118404	Alv_FOS4	DQ118404_Alv_FOS4	Archaea ; Euryarchaeota G2
DQ078753	Alv_FOS5	DQ078753_Alv_FOS5	Archaea ; Euryarchaeota G2
DQ156349	HF0070_39H11	HF70_39H11	Archaea ; Euryarchaeota G2
AY458629	EB750_01A01	EB750_01A01	putative Archaea
CH672415	*Agreia* sp. PHSC20c1	actino_PHSC20C1	Actinobacteria ; Actinobacteria ; Actinomycetales
CH672413	*Janibacter* sp. HTCC2649	Janibacter_HTCC2649	Actinobacteria ; Actinobacteria ; Actinomycetales
GU474880	HF0200_20K23	HF0200_20K23	Actinobacteria ; Actinobacteria ; Candidatus "Microthrix"
GU474860	HF0070_17F14	HF0070_17F14	Actinobacteria ; Actinobacteria ; Candidatus "Microthrix"
GU474887	HF4000_04C13	HF4000_04C13	Actinobacteria ; Actinobacteria ; Candidatus "Microthrix"
DQ295241	AntFos_39E11	AntFos_39E11	Bacteroidetes/Chlorobi group
DQ272742	*Bacteroidetes* clone SBI2_18 P41A3	DQ272742_Bacteroidetes	Bacteroidetes/Chlorobi group
AJ937771	Uncultured *Flavobacteriaceae* bacterium fosmid	AJ937771_Flavobacterial	Bacteroidetes/Chlorobi group ; Bacteroidetes ; Flavobacteria
AAPD01000000	*Flavobacteria* BBFL7	Flavobacteria_BBFL7	Bacteroidetes/Chlorobi group ; Bacteroidetes ; Flavobacteria
CH672373	*Croceibacter atlanticus* HTCC2559	C_atlanticus_HTCC2559	Bacteroidetes/Chlorobi group ; Bacteroidetes ; Flavobacteria
CH672391	Not yet validly described HTCC2170	Flavobacteriales_HTCC2170	Bacteroidetes/Chlorobi group ; Bacteroidetes ; Flavobacteria
CH672395	*Leeuwenhoekiella blandensis* MED217	Flavobacterium_MED217	Bacteroidetes/Chlorobi group ; Bacteroidetes ; Flavobacteria
CH724148	*Polaribacter irgensii* 23-P	Polaribacter_23_P	Bacteroidetes/Chlorobi group ; Bacteroidetes ; Flavobacteria
CH902588	*Polaribacter* sp. MED152	Polaribacter_MED152	Bacteroidetes/Chlorobi group ; Bacteroidetes ; Flavobacteria
CP001712	*Robiginitalea biformata* HTCC2501	Robiginitalea_HTCC2501	Bacteroidetes/Chlorobi group ; Bacteroidetes ; Flavobacteria
NZ_ABCM00000000	*Pedobacter* sp. BAL39	Pedobacter_BAL39	Bacteroidetes/Chlorobi group ; Bacteroidetes ; Sphingobacteria
GU474851	HF0010_19H17	HF0010_19H17	Bacteroidetes/Chlorobi group ; Bacteroidetes ; Sphingobacteria
GU474874	HF0130_33B19	HF0130_33B19	Bacteroidetes/Chlorobi group ; Bacteroidetes ; Sphingobacteriales ; OM273
AAMZ01000000	*Cellulophaga* sp. MED134	D_donghaensis_MED134	Bacteroidetes/Chlorobi group ; Bacteroidetes/Chlorobi group ; Flavobacteriales ; Dokdonia donghaensis MED134
DQ295240	AntFos_29B07	AntFos_29B07	Bacteroidetes/Chlorobi group ; putative Bacteroidetes
GU474939	EB080_L08E11	EB080_L08E11	CFB ; uncultivated Cytophaga
GU474838	HF0770_11D24	HF0770_11D24	Chlamydiae/Verrucomicrobia group ; Verrucomicrobia ; Uncultured Verrucomicrobia
GU474863	HF0070_30L02	HF0070_30L02	Chlamydiae/Verrucomicrobia group ; Verrucomicrobiales
GU474882	HF0200_39L05	HF0200_39L05	Chlamydiae/Verrucomicrobia group ; Verrucomicrobiales
GU474890	HF4000_13K17	HF4000_13K17	Chlamydiae/Verrucomicrobia group ; Verrucomicrobiales
GU474845	HF0010_05E02	HF0010_05E02	Chlamydiae/Verrucomicrobia group ; Verrucomicrobiales ; MB11C04 clade
GU474876	HF0200_06I16	HF0200_06I16	Chloroflexi ; Chloroflexi (class) ; Unclassified Chloroflexi
GU474878	HF0200_09I09	HF0200_09I09	Chloroflexi ; Chloroflexi (class) ; Unclassified Chloroflexi
GU474918	HF0500_03M05	HF0500_03M05	Chloroflexi ; Chloroflexi (class) ; Unclassified Chloroflexi
GU474924	HF0770_09E03	HF0770_09E03	Chloroflexi ; Chloroflexi (class) ; Unclassified Chloroflexi
GU474897	HF4000_28F02	HF4000_28F02	Chloroflexi ; Chloroflexi (class) ; Unclassified Chloroflexi
AAXW00000000	*Cyanothece* sp. CCY0110	Cyanothece_CCY0110	Cyanobacteria ; Chroococcales
CP000435	*Synechococcus* strain CC9311	Syn_CC9311	Cyanobacteria ; Chroococcales ; Synechococcus clade I
CP000110	*Synechococcus* strain CC9605	Syn_CC9605	Cyanobacteria ; Chroococcales ; Synechococcus clade II
BX548020	*Synechococcus sp. WH8102*	Syn_WH8102	Cyanobacteria ; Chroococcales ; Synechococcus clade III
AATZ00000000	*Synechococcus* sp. BL107	Syn_BL107	Cyanobacteria ; Chroococcales ; Synechococcus clade IV
CP000097	*Synechococcus* strain CC9902	Syn_CC9902	Cyanobacteria ; Chroococcales ; Synechococcus clade IV
CT971583	*Synechococcus* strain WH7803	Syn_WH7803	Cyanobacteria ; Chroococcales ; Synechococcus clade V
AAOK00000000	*Synechococcus* sp. WH7805	Syn_WH7805	Cyanobacteria ; Chroococcales ; Synechococcus clade VI
CT978603	*Synechococcus sp. RCC307*	Syn_RCC307	Cyanobacteria ; Chroococcales ; Synechococcus clade X
EF089389	HOT0_02H05	HOT0_02H05	Cyanobacteria ; Crocosphaera
EF089390	HOT0_07D09	HOT0_07D09	Cyanobacteria ; Crocosphaera
CP000552	*Prochlorococcus* sp. MIT9515	Pro_MIT_9515	Cyanobacteria ; Prochlorales ; Prochlorococcus HL clade ; low B/A clade I
BX548174	*Prochlorococcus* MED4 (aka CCMP1986, aka CCMP1378)	ProMED4	Cyanobacteria ; Prochlorales ; Prochlorococcus HL clade ; low B/A clade I
CP000111	*Prochlorococcus* str. MIT 9312	Pro_9312	Cyanobacteria ; Prochlorales ; Prochlorococcus HL clade ; low B/A clade II
CP000551	*Prochlorococcus* sp. AS9601	Pro_AS9601	Cyanobacteria ; Prochlorales ; Prochlorococcus HL clade ; low B/A clade II
AE017126	*Prochlorococcus* CCMP1375 = SS120	Pro_SS120_CCMP1375	Cyanobacteria ; Prochlorales ; Prochlorococcus LL clade
CP000095	*Prochlorococcus* sp. NATL2A	Pro_NATL2A	Cyanobacteria ; Prochlorales ; Prochlorococcus LL clade ; high B/A clade I
BX548175	*Prochlorococcus* str. MIT 9313	Pro_9313	Cyanobacteria ; Prochlorales ; Prochlorococcus LL clade ; high B/A clade IV
GU474867	HF0130_06E03	HF0130_06E03	EF100_108A04 cluster, which was previously in Agg47 by Suzuki et al 2004
GU474921	HF0500_16O16	HF0500_16O16	EF100_108A04 cluster, which was previously in Agg47 by Suzuki et al 2004
GU474881	HF0200_23L05	HF0200_23L05	Fibrobacteres/Acidobacteria group ; Acidobacteria ; Acidobacteria (class) ; Acidobacteriales
GU474896	HF4000_26D02	HF4000_26D02	Fibrobacteres/Acidobacteria group ; Acidobacteria ; Acidobacteria (class) ; Acidobacteriales
GU474926	HF0770_27F21	HF0770_27F21	Fibrobacteres/Acidobacteria group ; Acidobacteria ; Acidobacteria (class) ; Unclassified Acidobacteriales
DQ295242	AntFos_04D05	AntFos_04D05	Gemmatimonadetes ; Gemmatimonadales ; Gemmatimonadaceae ; Gemmatimonas
GU474865	HF0130_03D03	HF0130_03D03	Gemmatimonadetes ; Gemmatimonadetes ; Gemmatimonadales
DQ295238	AntFos_04E12	AntFos_04E12	Gram Positive High G + C
NZ_ABCK00000000	*Lentisphaera araneosa* HTCC2155	L_araneosa_HTCC2155	Lentisphaerae ; Lentisphaerales
EF089402	HF0010_49E08	HF10_49E08	Planctomycetales ; (by synteny with seq'd isolate, best BLAST hits 65.8%, and XylA phylogeny, McCarren & DeLong, 2007)
EF591885	INIKI_PLANKTO_6N14	Inikiplankto_6N14	Planctomycetes ; Pirellula-like?
EF591884	INIKI_PLANKTO_5H12	ORE200_05H12	Planctomycetes ; Pirellula-like?
CH672376	*Blastopirellula marina* DSM 3645T	B_marina_DSM_3645	Planctomycetes ; Planctomycetacia ; Planctomycetales
GU474923	HF0500_40D21	HF0500_40D21	Planctomycetes ; Planctomycetacia ; Planctomycetales ; Planctomycetaceae ; Planctomyces
BX119912	*Rhodopirellula baltica SH 1*	Rhodopirellula_SH_1	Planctomycetes ; Planctomycetacia ; Planctomycetales ; Planctomycetaceae ; Rhodopirellula ; Rhodopirellula baltica
EF107103	HF0010_45G01	HF10_45G01	Proteobacteria
EF089397	EB000_35D03	EB000_35D03	Proteobacteria
EF107099	EB000_49D07	EB000_49D07	Proteobacteria
EF100190	HF0010_19P19	HF10_19P19	Proteobacteria
EF089399	EB000_39H12	EB000_39H12	Proteobacteria
EF100191	HF0010_25F10	HF10_25F10	Proteobacteria
AY372455	HOT_02C01	HOT2C01	Proteobacteria ; Alphaproteobacteria
EF089398	EB000_39F01	EB000_39F01	Proteobacteria ; Alphaproteobacteria
EF107105	EB080_69G07	EB080_69G07	Proteobacteria ; Alphaproteobacteria
EF107102	HF0010_12C08	HF10_12C08	Proteobacteria ; Alphaproteobacteria
AE008920	EB000_29C02	EB000_29C02	Proteobacteria ; Alphaproteobacteria
GU474868	HF0130_06E21	HF0130_06E21	Proteobacteria ; Alphaproteobacteria ; D108 clade

AY458634	EB750_09G06	EB750_09G06	Proteobacteria ; Alphaproteobacteria ; putative SAR11
GU474873	HF0130_31E21	HF0130_31E21	Proteobacteria ; Alphaproteobacteria ; nuHF1 calde
GU474858	HF0070_14E07	HF0070_14E07	Proteobacteria ; Alphaproteobacteria ; OM75
GU474870	HF0130_20P23	HF0130_20P23	Proteobacteria ; Alphaproteobacteria ; OM75
CH724133	*Parvularcula bermudensis* HTCC2503	Parvularcula_HTCC2503	Proteobacteria ; Alphaproteobacteria ; Parvularculales
DS022272	*Fulvimarina pelagi* HTCC2506	F_pelagi_HTCC2506	Proteobacteria ; Alphaproteobacteria ; Rhizobiales
AAMY01000000	*Nitrobacter* sp. Nb 311A	Nitrobacter_Nb_311A	Proteobacteria ; Alphaproteobacteria ; Rhizobiales
GU474930	HF4000_48A13	HF4000_48A13	Proteobacteria ; Alphaproteobacteria ; Rhizobiales ; Parvibaculum
GU474898	HF4000_32B18	HF4000_32B18	Proteobacteria ; Alphaproteobacteria ; Rhizobiales ; SIMO CL-S30-58 clade
CP000264	*Jannaschia* CCS1	Jannaschia_CCS1	Proteobacteria ; Alphaproteobacteria ; Rhodobacterales
CH672414	*Loktanella vestfoldensis* SKA53	L_vestfoldensis_SKA53	Proteobacteria ; Alphaproteobacteria ; Rhodobacterales
CH672428	*Oceanicaulis alexandrii* HTCC2633	O_alexandrii_HTCC2633	Proteobacteria ; Alphaproteobacteria ; Rhodobacterales
CH724131	*Oceanicola batsensis* HTCC2597	O_batensis_HTCC2597	Proteobacteria ; Alphaproteobacteria ; Rhodobacterales
CH724107	*Oceanicola granulosus* HTCC2516	O_granulosus_HTCC2516	Proteobacteria ; Alphaproteobacteria ; Rhodobacterales
CH902578	*Rhodobacterales* HTCC2654 aka *Maritimibacter alkaliphilus* HTCC2654	Rhodobacterales_HTCC2654	Proteobacteria ; Alphaproteobacteria ; Rhodobacterales
AAYB01000000	*Roseobacter* sp. CCS2	Roseobacter_CCS2	Proteobacteria ; Alphaproteobacteria ; Rhodobacterales
CH902583	*Roseobacter* sp. MED193	Roseobacter_MED193	Proteobacteria ; Alphaproteobacteria ; Rhodobacterales
AAYC00000000	*Roseobacter* sp. SK209-2-6	Roseobacter_SK_209_2_6	Proteobacteria ; Alphaproteobacteria ; Rhodobacterales
CH902584	*Roseovarius* sp. 217	Roseovarius_217	Proteobacteria ; Alphaproteobacteria ; Rhodobacterales
DS022279	*Roseovarius* sp HTCC2601 aka *Pelagibaca bermudensis* HTCC2601	Roseovarius_HTCC2601	Proteobacteria ; Alphaproteobacteria ; Rhodobacterales
ABCL01000012	*Roseovarius* sp. TM1035	Roseovarius_TM1035	Proteobacteria ; Alphaproteobacteria ; Rhodobacterales
CP000031	*Silicibacter pomeroyi DSS-3*	S_pomeroyi_DSS-3	Proteobacteria ; Alphaproteobacteria ; Rhodobacterales
AAYA00000000	*Sagittula stellata* E37	S_stellata_E_37	Proteobacteria ; Alphaproteobacteria ; Rhodobacterales
CH959310	*Sulfitobacter* sp. EE-36	Sulfitobacter_EE36	Proteobacteria ; AlphaProteobacteria ; Rhodobacterales
CH959312	*Sulfitobacter* sp. NAS-14.1	Sulfitobacter_NAS14_1	Proteobacteria ; AlphaProteobacteria ; Rhodobacterales
AY458649	EB080_L28H02	EB080_L028H02	Proteobacteria ; Alphaproteobacteria ; putative Rhodobacterales
GU474931	EB000_36F02	EB000_36F02	Proteobacteria ; Alphaproteobacteria ; Rhodobacterales ; CHAB1-5
GU474942	EB080_L58F04	EB080_L58F04	Proteobacteria ; Alphaproteobacteria ; Rhodobacterales ; CHAB1-5
GU474937	EB080_L11F12	EB080_L11F12	Proteobacteria ; Alphaproteobacteria ; Rhodobacterales ; NAC11-7
GU474940	EB080_L27A02	EB080_L27A02	Proteobacteria ; Alphaproteobacteria ; Rhodobacterales ; NAC11-7
GU474941	EB080_L43F08	EB080_L43F08	Proteobacteria ; Alphaproteobacteria ; Rhodobacterales ; NAC11-7
GU474938	EB080_L06A09	EB080_L06A09	Proteobacteria ; Alphaproteobacteria ; Rhodobacterales ; NAC11-7
NZ_AATR00000000	*Rhodobacterales* HTCC2255	alpha_HTCC2255	Proteobacteria ; Alphaproteobacteria ; Rhodobacterales ; NAC11-7
AAXZ00000000	*Roseobacter* HTCC2150	Rhodobacterales_HTCC2150	Proteobacteria ; Alphaproteobacteria ; Rhodobacterales ; RCA clade
GU474843	HF0010_04M21	HF0010_04M21	Proteobacteria ; Alphaproteobacteria ; Rhodobacterales ; Roseobacter-like
GU474856	HF0070_10D05	HF0070_10D05	Proteobacteria ; Alphaproteobacteria ; Rhodobacterales ; Roseobacter-like
GU474886	HF4000_03E16	HF4000_03E16	Proteobacteria ; Alphaproteobacteria ; Rhodobacterales ; Roseobacter-like
GU474943	EB080_L84F03	EB080_L84F03	Proteobacteria ; Alphaproteobacteria ; Rhodobacterales ; Roseobacter
DQ295239	AntFos_24C04	AntFos_24C04	Proteobacteria ; Alphaproteobacteria ; Rhodobacterales ; Roseobacter
AE008921	EB000_60D04	EB060D04	Proteobacteria ; Alphaproteobacteria ; Rhodobacterales ; Roseobacter-like (by best BLAST hits)
AY671989	eBACred_25D05	EBACred_25D05	Proteobacteria ; Alphaproteobacteria ; Rhodobacterales ; Roseobacter-like bacteria puf/bcll
GU474905	HF0010_10C01	HF0010_10C01	Proteobacteria ; Alphaproteobacteria ; Rhodobacterales ; SAR102
GU474935	EB000_55B11	EB055B11	Proteobacteria ; Alphaproteobacteria ; Rhodobacter-like
GU474864	HF0070_31K06	HF0070_31K06	Proteobacteria ; Alphaproteobacteria ; Rhodospirillales
GU474875	HF0200_01O14	HF0200_01O14	Proteobacteria ; Alphaproteobacteria ; Rhodospirillales
GU474895	HF4000_24M03	HF4000_24M03	Proteobacteria ; Alphaproteobacteria ; Rhodospirillales
GU474947	EF100_102A06	EF100_102A06	Proteobacteria ; Alphaproteobacteria ; Rhodospirillales ; OM-75
EU795181	HF0070_02E07	HF0070_02E07	Proteobacteria ; Alphaproteobacteria ; Rickettsiales
AAPV00000000	*Pelagibacter ubique* HTCC1002	Pelagibacter_HTCC1002	Proteobacteria ; Alphaproteobacteria ; Rickettsiales ; SAR11 clade ; subgroup 1a
CP000084	*Pelagibacter ubique* HTCC1062	Pelagibacter_HTCC1062	Proteobacteria ; Alphaproteobacteria ; Rickettsiales ; SAR11 clade ; subgroup 1a
GU474840	HF4000_09C18	HF4000_384_009C18	Proteobacteria ; Alphaproteobacteria ; Rickettsiales ; SAR11 clade
GU474904	HF0010_09O16	HF0010_09O16	Proteobacteria ; Alphaproteobacteria ; Rickettsiales ; SAR11 clade
GU474927	HF0770_37D02	HF0770_37D02	Proteobacteria ; Alphaproteobacteria ; Rickettsiales ; SAR11 clade ; Pelagibacter
GU474900	HF4000_37C10	HF4000_37C10	Proteobacteria ; Alphaproteobacteria ; Rickettsiales ; SAR11 clade
AY458633	EB750_11E01	EBAC750_11E01	Proteobacteria ; Alphaproteobacteria ; Rickettsiales ; SAR11 clade ; SAR11 ; subgroup 2
AY458637	EB750_02H05	EB75002H05	Proteobacteria ; Alphaproteobacteria ; Rickettsiales ; SAR11 clade ; SAR11 ; subgroup 2
GU474946	EF100_94H03	EF100_94H03	Proteobacteria ; Alphaproteobacteria ; roots rhodovibrio
AY744399	eBACred_02C11	EBred_02C11	Proteobacteria ; Alphaproteobacteria ; SAR116 ; putative SAR116-I
GU474848	HF0010_13E22	HF0010_13E22	Proteobacteria ; Alphaproteobacteria ; SAR116
GU474859	HF0070_17D04	HF0070_17D04	Proteobacteria ; Alphaproteobacteria ; SAR116
GU474907	HF0070_05I22	HF0070_05I22	Proteobacteria ; Alphaproteobacteria ; SAR116
GU474910	HF0070_34A12	HF0070_34A12	Proteobacteria ; Alphaproteobacteria ; SAR116
GU474911	HF0070_34E11	HF0070_34E11	Proteobacteria ; Alphaproteobacteria ; SAR116
GU474934	EB000_46D07	EB000_46D07	Proteobacteria ; Alphaproteobacteria ; SAR116 ; SAR116-I
GU474933	EB000_37G09	EB000_37G09	Proteobacteria ; Alphaproteobacteria ; SAR116 ; SAR116-II
CP000157	*Erythrobacter litoralis* HTCC2594	E_litoralis_HTCC2594	Proteobacteria ; Alphaproteobacteria ; Sphingomonadales
GU474922	HF0500_24B12	HF0500_24B12	Proteobacteria ; Alphaproteobacteria ; Sphingomonadales ; Erythrobacteraceae ; Erythrobacter
GU474853	HF0010_30A23	HF0010_30A23	Proteobacteria ; Alphaproteobacteria ; T31_112 clade
EF089400	EB000_41B09	EB041B09	Proteobacteria ; Betaproteobacteria
CP000316	*Polaromonas* sp. JS666 - draft	Polaromonas_JS666	Proteobacteria ; Betaproteobacteria ; Burkholdariales ; Polaromonas
GU474839	HF4000_05M23	HF4000_05M23	Proteobacteria ; Betaproteobacteria ; Burkholderiales ; Delftia
AAUX00000000	Not yet validly described, OM43 clade HTCC2181	Methylophilales_HTCC2181	Proteobacteria ; Betaproteobacteria ; Methylophilales
AY458645	EB080_L12H07	EB80L12H07	Proteobacteria ; Betaproteobacteria ; Nitrosomonas
GU474866	HF0130_04F21	HF0130_04F21	Proteobacteria ; Betaproteobacteria ; OM156
AY458647	EB000_36A07	EB000_36A07	Proteobacteria ; Betaproteobacteria ; OM43
GU474901	HF0010_04H24	HF0010_04H24	Proteobacteria ; Betaproteobacteria ; Rhodocyclales ; Rhodocyclaceae ; Zoogloea
GU474906	HF0010_10I05	HF0010_10I05	Proteobacteria ; delta/epsilon subdivisions ; Deltaproteobacteria ; SAR324
GU474908	HF0070_07E19	HF0070_07E19	Proteobacteria ; delta/epsilon subdivisions ; Deltaproteobacteria ; SAR324
GU474909	HF0070_15B21	HF0070_15B21	Proteobacteria ; delta/epsilon subdivisions ; Deltaproteobacteria ; SAR324
GU474912	HF0130_05G09	HF0130_05G09	Proteobacteria ; delta/epsilon subdivisions ; Deltaproteobacteria ; SAR324
GU474913	HF0130_20J24	HF0130_20J24	Proteobacteria ; delta/epsilon subdivisions ; Deltaproteobacteria ; SAR324
GU474914	HF0200_14D13	HF0200_14D13	Proteobacteria ; delta/epsilon subdivisions ; Deltaproteobacteria ; SAR324
GU474915	HF0200_39L23	HF0200_39L23	Proteobacteria ; delta/epsilon subdivisions ; Deltaproteobacteria ; SAR324
GU474883	HF0200_39N20	HF0200_39N20	Proteobacteria ; delta/epsilon subdivisions ; Deltaproteobacteria ; SAR324
GU474917	HF0500_03A04	HF0500_03A04	Proteobacteria ; delta/epsilon subdivisions ; Deltaproteobacteria ; SAR324
GU474837	HF0770_09N20	HF0770_09N20	Proteobacteria ; delta/epsilon subdivisions ; Deltaproteobacteria ; SAR324
GU474928	HF0770_45N15	HF0770_45N15	Proteobacteria ; delta/epsilon subdivisions ; Deltaproteobacteria ; SAR324
GU474869	HF0130_19C20	HF0130_19C20	Proteobacteria ; delta/epsilon subdivisions ; Deltaproteobacteria ; SAR324
GU474879	HF0200_19J16	HF0200_19J16	Proteobacteria ; delta/epsilon subdivisions ; Deltaproteobacteria ; SAR324
GU474888	HF4000_08N17	HF4000_08N17	Proteobacteria ; delta/epsilon subdivisions ; Deltaproteobacteria ; SAR324 cluster ; ctg_NISA008 clade
GU474862	HF0070_30B07	HF0070_30B07	Proteobacteria ; delta/epsilon subdivisions ; Deltaproteobacteria ; SAR324 cluster ; SAR276 clade
GU474903	HF0010_08B07	HF0010_08B07	Proteobacteria ; delta/epsilon subdivisions ; Deltaproteobacteria
AY458630	EB750_01B07	EB750_01B07	Proteobacteria ; delta/epsilon subdivisions ; Deltaproteobacteria

GU474877	HF0200_07G10	HF0200_07G10	Proteobacteria ; delta/epsilon subdivisions; Deltaproteobacteria ; Desulfobacterales ; Nitrospina-like
GU474836	HF0130_12L15	HF0130_12L15	Proteobacteria ; delta/epsilon subdivisions; Deltaproteobacteria ; Myxococcales ; E48F11cD clade
GU474842	HF0010_01J10	HF0010_01J10	Proteobacteria ; delta/epsilon subdivisions; Deltaproteobacteria ; OM27
GU474857	HF0070_10I02	HF0070_10I02	Proteobacteria ; delta/epsilon subdivisions; Deltaproteobacteria ; OM27
AY458631	EB750_03B02	EB750_03B02	Proteobacteria ; delta/epsilon subdivisions; Deltaproteobacteria ; SAR406
GU474916	HF0500_01L02	HF0500_01L02	Proteobacteria ; delta/epsilon subdivisions; Deltaproteobacteria ; SAR406 cluster
GU474850	HF0010_18O13	HF0010_18O13	Proteobacteria ; delta/epsilon subdivisions; Deltaproteobacteria ; SAR406 cluster ; A313008 clade
GU474892	HF4000_22B16	HF4000_22B16	Proteobacteria ; delta/epsilon subdivisions; Deltaproteobacteria ; SAR406 cluster ; ESP200-K10-15 clade
DQ267495	DeepAnt_1F12	DQ267495_DeepAnt_1F12	Proteobacteria ; Deltaproteobacteria ;
DQ267496	DeepAnt_32C6	DQ267496_DeepAnt_32C6	Proteobacteria ; Deltaproteobacteria ;
EF106972	EB080_L20F04	EB080_L20F04	Proteobacteria ; Deltaproteobacteria ; Nitrospinaceae
AE008919	EB000_65D09	EB065D09	Proteobacteria ; Gammaproteobacteria
AY458650	EB750_10A10	EB75010A10	Proteobacteria ; Gammaproteobacteria
AY458636	EB750_10B11	EB750_10B11	Proteobacteria ; Gammaproteobacteria
NZ_AAVT00000000	Not yet validly described HTCC2143	proteo_HTCC2143	Proteobacteria ; Gammaproteobacteria
NZ_AAOE00000000	*Reinekea* sp. MED297	Reinekea_MED297	Proteobacteria ; Gammaproteobacteria
EF107106	HF0130_81H07	HF130_81H07	Proteobacteria ; Gammaproteobacteria
GU474833	HF0010_16H03	HF10_16H03	Proteobacteria ; Gammaproteobacteria
EF107100	EB000_50A10	EB000_50A10	Proteobacteria ; Gammaproteobacteria
GU474945	EF100_93H11	EF100_93H11	Proteobacteria ; Gammaproteobacteria ; AGG47
GU474884	HF0200_40H22	HF0200_40H22	Proteobacteria ; Gammaproteobacteria ; AGG47
AAVS00000000	Not yet validly described TW-7	Alteromonadales_TW_7	Proteobacteria ; Gammaproteobacteria ; Alteromonadales
NC_011138	*Alteromonas macleodii* Deep ecotype	A_macleodii_Deep	Proteobacteria ; Gammaproteobacteria ; Alteromonadales
AAOH01000000	*Pseudoalteromonas tunicata* D2	Pseudo_tunicata_D2	Proteobacteria ; Gammaproteobacteria ; Alteromonadales
GU474929	HF4000_16C08	HF4000_16C08	Proteobacteria ; Gammaproteobacteria ; Alteromonadales
AE017340	*Idiomarina loihiensis L2TR*	I_loihiensis_L2TR	Proteobacteria ; Gammaproteobacteria ; Alteromonadales
DQ295237	AntFos_04D03	AntFos_04D03	Proteobacteria ; Gammaproteobacteria ; ArRCTIC96B-16
AY458646	EB080_L31E09	EB080_L31E09	Proteobacteria ; Gammaproteobacteria ; ARCTIC96BD-19 clade
CH672427	*Nitrococcus mobilis* Nb 231	N_mobilis_Nb_2311_1	Proteobacteria ; Gammaproteobacteria ; Chromatiales
GU474885	HF0200_41F04	HF0200_41F04	Proteobacteria ; Gammaproteobacteria ; Chromatiales ; Bivalve endosymbiont clade
GU474925	HF0770_27E13	HF0770_27E13	Proteobacteria ; Gammaproteobacteria ; DHB-2 Cluster
GU474936	EB000_65A11	EB000_65A11	Proteobacteria ; Gammaproteobacteria ; EB000_65A11 clade
GU474841	HF0010_01E20	HF0010_01E20	Proteobacteria ; Gammaproteobacteria ; K189A clade
GU474849	HF0010_16J05	HF0010_16J05	Proteobacteria ; Gammaproteobacteria ; K189A clade
GU474855	HF0070_08D07	HF0070_08D07	Proteobacteria ; Gammaproteobacteria ; K189A clade
GU474932	EB000_37F04	EB000_37F04	Proteobacteria ; Gammaproteobacteria ; KTc1119 clade
GU474899	HF4000_36I10	HF4000_36I10	Proteobacteria ; Gammaproteobacteria ; NEP4 cluster (close to OM60 cluster)
NZ_AANE00000000	*Marinomonas* sp. MED121	Marinomonas_MED121	Proteobacteria ; Gammaproteobacteria ; Oceanospirillales
GU474861	HF0070_21F08	HF0070_21F08	Proteobacteria ; Gammaproteobacteria ; Oceanospirillales ; Alcanivorax-like
GU474889	HF4000_13G19	HF4000_13G19	Proteobacteria ; Gammaproteobacteria ; Oceanospirillales ; Alcanivorax-like
GU474872	HF0130_25G24	HF0130_25G24	Proteobacteria ; Gammaproteobacteria ; Oceanospirillales ; OM182 clade
GU474894	HF4000_23O15	HF4000_23O15	Proteobacteria ; Gammaproteobacteria ; Oceanospirillales ; OM182 clade
AY458641	EB080_L32B05	EB080_L32B05	Proteobacteria ; Gammaproteobacteria ; OM60
GU474844	HF0010_05D02	HF0010_05D02	Proteobacteria ; Gammaproteobacteria ; OM60
GU474835	HF0130_01F24	HF0130_01F24	Proteobacteria ; Gammaproteobacteria ; OM60
AAOA00000000	*Congregibacter litoralis* KT 71	gamma_KT_71	Proteobacteria ; Gammaproteobacteria ; OM60 clade
AAVV00000000	OM60 clade, HTCC2080	proteo_HTCC2080	Proteobacteria ; Gammaproteobacteria ; OMG, OM60 clade
GU474920	HF0500_12O04	HF0500_12O04	Proteobacteria ; Gammaproteobacteria ; Pseudomonadales ; Pseudomonas
AY458632	EB750_02H09	EB750_02H09	Proteobacteria ; Gammaproteobacteria ; SAR156
GU474891	HF4000_19M20	HF4000_19M20	Proteobacteria ; Gammaproteobacteria ; SAR156 cluster ; EB750_02H09 clade
GU474846	HF0010_11B23	HF0010_11B23	Proteobacteria ; Gammaproteobacteria ; SAR86 cluster ; CHAB-I-7 clade
GU474871	HF0130_23I23	HF0130_23I23	Proteobacteria ; Gammaproteobacteria ; SAR86 cluster ; CHAB-I-7 clade
GU474854	HF0070_03O15	HF70_03O15	Proteobacteria ; Gammaproteobacteria ; SAR86 cluster ; CHAB-I-7 clade
GU474847	HF0010_11K06	HF0010_11K06	Proteobacteria ; Gammaproteobacteria ; SAR86 cluster ; SAR89 clade
AY552545	eBACred_20E09	EBACred_20E09	Proteobacteria ; Gammaproteobacteria ; SAR86-I
AY619685	HOT_04E07	HOT4E07	Proteobacteria ; Gammaproteobacteria ; SAR86-I
AF279106	EB000_31A08	EB0_31A08	Proteobacteria ; Gammaproteobacteria ; SAR86-II
AY372454	EB000_45B06	EB000_45B06	Proteobacteria ; Gammaproteobacteria ; SAR86-II
EF107104	EB080_02D08	EB080_02D08	Proteobacteria ; Gammaproteobacteria ; SAR-86-II
GU474944	EBAC_27G05	EBAC_27G05	Proteobacteria ; Gammaproteobacteria ; SAR86-III
GU474902	HF0010_05E14	HF0010_05E14	Proteobacteria ; Gammaproteobacteria ; SAR92
GU474893	HF4000_23L14	HF4000_23L14	Proteobacteria ; Gammaproteobacteria ; Thiotricales ; ZD0405 clade
AY744396	eBACred_07D11	EBACred_07D11	Proteobacteria ; Gammaproteobacteria ; putative Vibrionales
CP000020	*Vibrio fischeri* ES114	Vibrio_fischeri_ES114	Proteobacteria ; Gammaproteobacteria ; Vibrionales ; Vibrio
CH724174	*Vibrio splendidus* 12B01	V_splendidus_12B01	Proteobacteria ; Gammaproteobacteria ; Vibrionales ; Vibrio ; splendidus group
CH902608	*Vibrio* sp. MED222	Vibrio_MED222	Proteobacteria ; Gammaproteobacteria ; Vibrionales ; Vibrio ; splendidus group
AAZW00000000	*Vibrio* sp. SWAT-3	Vibrio_SWAT_3	Proteobacteria ; Gammaproteobacteria ; Vibrionales ; Vibrio ; splendidus group
GU474852	HF0010_22E23	HF0010_22E23	Proteobacteria ; Gammaproteobacteria ; Vibrionales ; Vibrio/Photobacterium-like
CR354531	*Photobacterium profundum SS9*	Photobacterium_SS9	Proteobacteria ; Gammaproteobacteria ; Vibrionales ; Photobacterium profundum
GU474949	EB750_07C09	EB750_07C09	Proteobacteria ; Gammaproteobacteria ; ZD0408
GU474948	EB080_L93H08	EB080_L93H08	Proteobacteria ; Gammaproteobacteria ; ZDO417
DQ068067	MED13K09	DQ068067_MED13K09	Proteobacteria ; putative Gammaproteobacteria
AY458640	EB000_65D02	EB000_65D02	Proteobacteria
AY458643	EB000_47H08	EB000_47H08	Proteobacteria
AY372453	ANT_32C12	ANT32C12	Proteobacteria
AY372452	ANT_8C10	ANT8C10	Proteobacteria
GU474919	HF0500_06B09	HF0500_06B09	Spirochaetes ; Spirochaetales ; Rhizobiales ; Spirochaeta
GU474834	HF0070_11A08	L35766_PhN	unkown ; DMSP degradation, phosphonate degradation
DQ068068	RED17H08	DQ068068_RED17H08	unknown ; PR in alphaproteobacterial-PR clade
DQ065755	66A03	DQ065755_66A03	unknown ; PR in alphaproteobacterial-PR clade
DQ088847	MedeBAC46A06	DQ088847_MedeBAC46A06	unknown ; PR in alphaproteobacterial-PR clade
DQ073796	MedeBAC82F10	DQ073796_MedeBAC82F10	unknown ; PR in alphaproteobacterial-PR clade
DQ077553	MedeBAC35C06	DQ077553_MedeBAC35C06	unknown ; PR in alphaproteobacterial-PR clade
DQ077554	MedeBAC49C08	DQ077554_MedeBAC49C08	unknown ; PR in alphaproteobacterial-PR clade

Table S2. Array targets summarized by phylogenetic clade

Note: bolded lines correspond to clades used in Table S3

Phylogenetic clade	# of Clones Targeted	# of Genomes Targeted
Gammaproteobacteria	*46*	*15*
Vibrionales	**2**	**5**
SAR92	**1**	**0**
Alteromonadales	1	4
Arctic96B-16	**1**	**0**
Oceanospirillales	4	1
SAR86-I	**2**	**0**
SAR86-II	**3**	**0**
SAR86-III	**1**	**0**
CHAB-I-7	3	0
SAR89	1	0
SAR156	**2**	**0**
EB000-65A11	**1**	**0**
KTc1119	**1**	**0**
OM60	**3**	**2**
Agg47	**2**	**0**
Arctic96BD-19	**1**	**0**
ZD0417	**1**	**0**
ZD0408	**1**	**0**
Chromatiales	1	1
DHB-2	1	0
K189A	3	0
NEP4	1	0
Pseudomonadales	1	0
ZD0405	1	0
unclassified	7	2
Betaproteobacteria	*6*	*2*
OM43	**1**	**0**
Burkholdariales	1	1
Methylophilales	0	1
OM156	1	0
Rhodocyclales	1	0
Nitrosomonas	**1**	**0**
unclassified	1	0
Alphaproteobacteria	*50*	*23*
NAC11-7	**4**	**1**
CHAB-1-5	**2**	**0**
other Rhodobacterales	10	17
OM75	**3**	**0**
other Rhodospirillales	3	0
EF100-94H03	**1**	**0**
SAR116	**8**	**0**
Pelagibacter (SAR11)	**7**	**2**
other Rickettsiales	1	0
Sphingomonadales	1	1
T31_112	1	0
Rhizobiales	2	2
D108	1	0
nuHF1	1	0
unclassified	5	0
Deltaproteobacteria	*28*	*0*
SAR324	**15**	**0**
SAR406	**4**	**0**
OM27	2	0
Myxococcales	1	0
Nitrospina	**2**	**0**
unclassified	4	0
Unclassified Proteobacteria	*10*	*0*
Spirochaetes	*1*	*0*
Planctomycetes	*4*	*2*
Gemmatimonadetes	*2*	*0*
Lentisphaerae	*0*	*1*
Gram Positive High G + C	*1*	*0*
Bacteroidetes	***6***	***9***
Acidobacteria	*3*	*0*
EF100_108A04 cluster	*2*	*0*
Chloroflexi	*5*	*0*
Cytophaga	*1*	*0*
Verrucomicrobiales	***5***	***0***
Cyanobacteria	*2*	*16*
Prochlorochococcus	0	7
Synechococcus	**0**	**8**
Crocosphaera	2	0
Cyanothece	0	1
Marine Actinobacteria	***3***	***2***
Marine Crenarchaeota	***2***	***0***
Marine Euryarchaeota	***12***	***0***
Unclassified Archaea	*1*	*0*
Unidentified	*7*	*0*

Table S3. Comparison of array with other broad taxonomic surveys of Monterey Bay.

Data type	**Relative abundance by genome proxy array[1]**			**Percentages of each clade in large-insert clone libraries[2]**				**% of total community by QPCR[3]**
Location in Monterey Bay	Station M1 (36°45.50N 122°02.10W)			Station M2 (36.78N, 122.48W)	Sation M1 (36°45.50N 122°02.10W)	Station M1 (36°45.50N 122°02.10W)	(36°41.1319 N 122°02.3727 W)	MB upwelling plume
Date of sample(s)	2000-2004			Mar 1999	July 1999	Feb 2002	Apr 2000	Apr 2000
Depth of sample(s)	0m	30m	200m	3m	80m	100m	750m	5m
Phylogenetic clade								
Gammaproteobacteria	++++ [a]	++++	++++	*16.9*	*34.5*	*27.5*	*22.2*	
Vibrio	– [b]	–	–	–	–	3.1	–	
SAR92	–	–	–	–	1.4	–	–	
Pseudoalteromonas	–	–	–	–	0.3	–	–	
Arctic96B-16	–	–	–	–	1.4	–	–	
Marinomonas	–	–	–	–	0.3	–	–	
SAR86-I	–	–	–	2.8	–	–	–	
SAR86-II	++++	++++	++	1.4	2	–	–	~0.5-6%
SAR86III	+++	+++	+	1.4	3.4	1.6	–	
SAR156	–	–	++++	–	3.7	–	7.4	
EB000-65A11	++	+++	++	1.4	0.7	4.7	–	
KTc1119	+	–	–	1.4	–	0.8	–	
OM60	+	–	–	2.8	0.7	1.6	–	
Agg47	–	–	–	–	–	6.3	–	
Arctic96BD-19	+	+++	++++	5.6	19.9	5.5	7.4	
ZD0417	+	+	++++	–	0.3	–	–	
ZD0408	–	–	–	–	–	–	3.7	
Betaproteobacteria	+++	+++	+	1.4	3	–	–	
OM43	+++	+++	–	1.4	2.7	–	–	
Nitrosomonas	–	+	+	–	0.3	–	–	
Alphaproteobacteria	++++	++++	++++	*50.7*	*39.2*	*11.7*	*40.7*	
NAC11-7	++++	++++	++++	21.1	23.6	–	–	
CHAB-1-5	+++	+++	++++	5.6	5.7	0.8	–	~10-38%
Other Roseobacter clades	++	+	–	7	6.4	2.3	–	
EF100-94H03	–	–	–	–	–	0.8	3.7	
SAR116	+++	+++	–	11.3	1.4	0.8	–	
Pelagibacter (SAR11)	+++	+++	+++	5.6	2	7	37	~2-18%
OM75	–	–	–	–	–	2.3	–	
Deltaproteobacteria	–	+	++++	–	*0.7*	–	*7.4*	
SAR324	–	+	++++	–	0.3	–	7.4	
Nitrospina	–	+	++	–	0.3	–	–	
Bacteroidetes	++++	++	+	*8.5*	*1.4*	–	–	*~2-6%[4]*
SAR406 (Fibrobacter)	–	–	+++	–	–	–	*3.7*	
Verrucomicrobiales	+	–	++	–	*2*	*4.7*	*3.7*	
Cyanobacteria	++++	+++	–	*12.7*	*0.3*	*1.6*	–	*~0-.25%*
Synechococcus	+++	+++	–	1.4	0.3	1.6	–	
Marine Actinobacteria	+	+	+	–	*3*	*1.6*	–	
Marine Crenarchaeota	–	–	+++	–	–	*3.9*	–	
Marine Euryarchaeota	++	++	++++	*1.4*	–	*3.9*	–	
Unidentified	++++	++++	++++	*8.5*	*15.5*	*44.5*	*22.2*	

1. Data from this paper. 2. Data from Suzuki et al., 2004. 3. Data from Suzuki et al., 2001b. 4. Method targeted only the Cytophagales. a. "–" indicates none detected. b. + signs indicate at least 1 genotype within clade was present; ++++ = at least one genotype was present in 90-100% of samples, as denoted by the term "consistent" in the text, +++ = 50-90% of samples, as denoted by the term "frequent" in the text, ++ = 25-50% of samples, + = 0-25% of samples. Shaded cells indicate phylogenetic group not targeted by Suzuki et al., 2001; clades documented by Suzuki et al., 2004, but not targeted by any genotypes on the array were omitted from this table. Note that the array does not comprehensively target the genotypic space within each clade, unlike the 16S-screening and FISH-based methods; a negative "–" by the array indicates only that the targeted genotypes were absent not that the entire clade was assayed but absent.