Title

A microarray for assessing transcription from pelagic marine microbial taxa

Journal

The ISME Journal: Multidisciplinary Journal of Microbial Ecology, 8(7)

ISSN

1751-7362

Authors

Shilova, Irina N

Robidart, Julie C

James Tripp, H

et al.

Publication Date

2014-07-01

DOI

10.1038/ismej.2014.1

Peer reviewed

A microarray for assessing transcription from pelagic marine microbial taxa.

Irina N. Shilova[1*], Julie C. Robidart[1], H. James Tripp[2], Kendra Turk-Kubo[1], Boris Wawrik[3], Anton F. Post[4], Anne W. Thompson[1], Bess Ward[5], James T. Hollibaugh[6], Andy Millard[7], Martin Ostrowski[7], David Scanlan[7], Ryan W. Paerl[8], Rhona Stuart[8], and Jonathan P. Zehr[1]

[1]University of California Santa Cruz, Santa Cruz, CA, USA; [2]DOE Joint Genome Institute, Walnut Creek, CA, USA; [3]University of Oklahoma, USA; [4]Marine Biological Laboratory, Woods Hole, MA, USA; [5]Princeton University, Princeton, NJ, USA; [6]University of Georgia, Athens, GA, USA; [7]University of Warwick, Coventry, UK; [8]University of California San Diego, San Diego, CA, USA.

* Corresponding author: Department of Ocean Sciences, University of California Santa Cruz, 1156 High St., Santa Cruz, CA, USA; phone +1 (831) 459-5846; email: iirina@ucsc.edu.

SUBJECT CATEGORY: Integrated genomics and post-genomics approaches in microbial ecology

SHORT TITLE: MicroTOOLs microarray for marine microbial taxa

KEYWORDS: marine, microbial, microarray, transcription, molecular

Abstract

Metagenomic approaches have revealed unprecedented genetic diversity within microbial communities across vast expanses of the world's oceans. Linking this genetic diversity with key metabolic and cellular activities of microbial assemblages is a fundamental challenge. Here, we report on a collaborative effort to design MicroTOOLs (Microbiological Targets for Ocean Observing Laboratories), a high-density oligonucleotide microarray that targets functional genes of diverse taxa in pelagic and coastal marine microbial communities. MicroTOOLs integrates nucleotide sequence information from disparate data types: genomes, PCR-amplicons, metagenomes and metatranscriptomes. It targets 19,400 unique sequences over 145 different genes that are relevant to stress responses and microbial metabolism across the three domains of life and viruses. MicroTOOLs was used in a proof-of-concept experiment that compared the functional responses of microbial communities following Fe and P enrichments of surface water samples from the North Pacific Subtropical Gyre. We detected transcription of 68% of the gene targets across major taxonomic groups, and the pattern of transcription indicated relief from Fe limitation and transition to N limitation in some taxa. *Prochlorococcus* (eHLI), *Synechococcus* (sub-cluster 5.3) and Alphaproteobacteria SAR11 clade (HIMB59) showed the strongest responses to the Fe enrichment. In addition, members of uncharacterized lineages also responded. The MicroTOOLs microarray provides a robust tool for comprehensive characterization of major functional groups of microbes in the open ocean, and the design can be easily amended for specific environments and research questions.

Introduction

Marine microbial communities are complex, composed of diverse groups of Bacteria, Archaea, Eukaryotes and viruses. Molecular techniques frequently used in marine microbial ecology have shown strain-specific differences in genetic capabilities and transcriptional responses among the most abundant representatives of microbial communities (Fuhrman et al., 2006; Giovannoni and Vergin, 2012). Clades of *Prochlorococcus* in the North Atlantic have more phosphorus-acquisition strategies than clades in the North Pacific, as an adaptation to chronic phosphate limitation (Coleman and Chisholm, 2010). Coastal clades of *Synechococcus* have higher number of regulatory systems and the use for metals than open ocean clades, the latter being adapted to relatively constant oligotrophic conditions (Palenik *et al.,* 2006). To link ocean processes to microbial metabolism and to build better models for predicting responses to future ocean states (Azam and Malfatti, 2007), in light of this strain-level heterogeneity, new research tools are needed that assess individual and microbial community responses.

Microarray technology can complement more commonly used molecular techniques, such as PCR- and next generation sequencing, to provide cost-effective high throughput gene and transcript detection from several organisms in a single sample. Microarrays have the advantages of sample replication, standardization and robust interpretations of strain-level variation in functional gene transcriptional patterns and lend themselves to better comparative quantification of specific genes and transcripts, especially in rare organisms.

Phylogenetic and functional microarrays have been developed and used for identification of microorganisms and their activity in diverse environments. The most

comprehensive microbial functional microarray to date, the GeoChip 4.0, targets >10,000 sequences represented by 150 genes mainly from soil microbial communities and reduction-oxidation processes (Bai *et al.*, 2012; He *et al.*, 2010; He *et al.*, 2007; Zhou *et al*., 2013). In addition to contaminated soils, acid mine drainage sites, and Antarctic soils (Mason *et al.*, 2010; Xie *et al.*, 2011; Yergeau *et al.*, 2007), the GeoChip has been applied to detect microbial DNA and RNA in the marine environment (Lu *et al.*, 2012; Wawrik *et al.*, 2012). Aside from the PhyloChip, which targets 16S rRNA genes (Brodie *et al.*, 2006), existing microarrays target a specific genus (Rinta-Kanto *et al.*, 2011), a particular process (Abell *et al*., 2012; Bouskill *et al.*, 2011; Bulow *et al.*, 2008; Moisander *et al.*, 2007; Moisander *et al.*, 2006; Tiquia *et al.*, 2004; Ward *et al.*, 2007; Wu *et al.*, 2008), or ecosystem (Rich *et al.*, 2008; Rich *et al.*, 2011; Smith *et al.*, 2010). A comprehensive microarray that targets functional genes across diverse pelagic marine microbial communities has not yet been reported.

Various strategies have been used to overcome the lack of *a priori* knowledge of genomic sequences in target communities, which is a major limitation in the design of an environmental microarray. The most common strategy is to search public nucleotide sequence databases (such as NCBI) using key words (He *et al.*, 2010; He *et al.*, 2007; Rhee *et al.*, 2004; Smith *et al.*, 2010; Wu *et al.*, 2008). The resulting datasets, however, typically do not resemble the natural diversity of target genes, a problem that is exacerbated in undersampled environments. A second common strategy is to PCR-amplify a gene of interest and then either spot the products on a glass surface (Wu *et al.*, 2001) or use the derived sequence data in the design of oligonucleotide probes. While this technique results in a fingerprint for an uncharacterized community and is especially

valuable for rare targets (Bulow *et al.*, 2008; Moisander *et al.*, 2006; Taroncher-Oldenburg *et al.*, 2003; Ward *et al.*, 2007), the time-consuming cloning process makes this approach suitable for studying only a limited number of genes. The large nucleotide databases obtained with next generation sequencing (metagenomics) present an additional opportunity to access a cross-section of the diversity of a marker gene in natural populations.

Here, we report the design and application of a high-density oligonucleotide microarray, referred to as the MicroTOOLs microarray, that targets 19,400 sequences across 145 genes associated with open ocean and coastal microbial communities. The microarray is based on existing data from genomes, metagenomes, metatranscriptomes, and PCR-based assays and targets functional genes responsible for biogeochemical cycling and stress responses characteristic of the oceanic photic zone. As a proof-of-concept experiment, we analyzed microbial community responses to nutrient enrichments of inorganic phosphate (P) or ferric iron (Fe) from the oligotrophic Station (Stn.) ALOHA in the North Pacific Subtropical Gyre (NPSG). The physics, chemistry and biology of this region is well-characterized (Karl and Lukas, 1996; White *et al*., 2007), and is its microbial metagenome is well represented in the nucleotide sequence databases (DeLong *et al*., 2006; Frias-Lopez *et al.,* 2008; Hewson *et al.,* 2010). In the North Pacific, where N_2 fixation is a major supply of nitrogen (Karl, 1997), phosphorus and iron can be the limiting nutrients for N_2 fixation and primary production (Karl *et al.,* 2001; Moore *et al*., 2006; Grabowski *et al*., 2008; Karl and Letelier, 2008; Watkins-Brandt *et al*., 2011). We hypothesized that if microbial taxa were Fe-limited, Fe addition would result in decreased transcription of iron stress genes and increased transcription of

genes for energy, carbon, and nitrogen metabolism. If microbial taxa were P limited, P addition would result in decreased transcription of P stress genes and increased transcription of genes for energy metabolism, DNA replication, and cell division. However, we expected to see heterogeneous transcriptional responses from individual taxa due to differences in their genomic capabilities, nutrient requirements, and life strategies (Tolonen *et al.,* 2006; Dupont *et al.,* 2008; Ilikchyan et al 2009; Tetu *et al.,* 2009; Stuart *et al*., 2009; Kamennaya and Post, 2010; Thompson *et al*., 2011).

Methods

Design of the MicroTOOLs microarray

Target genes for oligonucleotide probe design were selected based on existing knowledge of gene-markers tracking microorganism interactions with their environment (e.g. Chen *et al.*, 2004; Dyhrman and Haley, 2006; Fuller *et al.*, 2005; Holtzendorff *et al.*, 2002; Kamennaya and Post, 2011; Lindell and Post, 2005; Mosier and Francis, 2011; Orchard *et al.*, 2009; Paerl *et al.*, 2011; Sebastian and Ammerman, 2009; Webb *et al.*, 2001; Zehr *et al.*, 2007). Several genes for hypothetical proteins that are differentially expressed in response to specific stimuli in cultured marine microorganisms were also included (Martiny *et al.*, 2009; Scanlan *et al.*, 1996; Shi *et al.*, 2009; Tetu *et al.*, 2009; Thompson *et al.*, 2011). A total of 145 genes provided molecular markers for metabolic and cellular processes (Table 1).

To obtain gene probes that adequately represent environmental nucleotide diversity, we searched all marine metagenomic and metatranscriptomic databases along with sequences from clone libraries for known genes. For metagenomic and metatranscriptomic searches, a seed amino acid sequence dataset was built for selected

taxa (Supplemental Material) for each targeted gene. This seed dataset was used for a TBLASTN query against "All Sanger reads" and "All 454 reads" in the Community Cyberinfrastructure for Advanced Microbial Ecology Research and Analysis (CAMERA, http://camera.calit2.net/, Sun *et al.*, 2011) with loose criteria: an E value cutoff of 1.0E-03 and up to 1,000 hits per query. All TBLASTN hits with lengths greater than 150 nucleotides (nt) were used as queries in reciprocal BLASTX in CAMERA to confirm gene annotation. For this analysis, an E value cutoff of 1E-10 and 60% percent identity over a minimum 40 amino acid (aa) alignment were used. Subsequent BLASTN searches in CAMERA were used to retrieve corresponding taxonomic information. Sequences with >85% identity over 100 nt alignment length to targeted marine microorganisms were selected. Taxonomic affiliation means that the target sequence had the highest similarity to a specific organism by BLASTN search against the 'nr' database. Thus, the accuracy of affiliation to an organism depended on gene conservation and their representation in sequence databases. NimbleGen technology allows 5% nucleotide mismatch in the whole probe region, thus sequences within a range of 95-100% nucleotide identity to the target gene were detected. We use the phrase 'organism-like genotype' to refer to target gene affiliation if nucleotide identity of the target gene to this organism was less than 95%. After reciprocal BLAST, sequences originating from metagenomic libraries that contained non-transcribed regions were trimmed at 5'- and 3'-ends of the ORF region. Custom Java applications and R scripts were developed to filter all BLAST results and to trim the ends of sequences, and are available upon request. Additionally, target sequences were added that derived from the clone libraries of genes (Supplementary Material) and from genomes of marine microorganisms contained in NCBI Genbank. Combined

sequences were clustered using CD-HIT-EST (Huang *et al.*, 2010; Li and Godzik, 2006) at 95% nt similarity. The longest representative sequence from each cluster was selected as the target sequence for oligonucleotide probe design. Probe design was performed at Roche NimbleGen (Madison, WI), and six probes of 60 nt length were designed for each target. Random oligonucleotide probes were included in addition to standard control and alignment NimbleGen probes. All oligonucleotide probes were tested *in silico* for possible cross-hybridization (Supplementary Material). Prior to the MicroTOOLs microarray design, we tested the specificity of oligonucleotide microarray technology in detecting transcripts from a mixed community and from environmental samples. The results showed cross-hybridization for highly conserved genes (such as the photosystem II *psbA* gene across cyanobacteria) when nucleotide sequences were < 5% dissimilar (Supplementary Material, Figure S1). The final design of the MicroTOOLs microarray comprised of ca. 116,000 experimental and 19,000 control probes with one replication synthesized on a twelve-plex 12X135K NimbleGen array. The platform is available at NCBI Gene Expression Omnibus (www.ncbi.nlm.nih.gov/geo/) under accession number GPL16706.

Nutrient enrichment incubation

An incubation experiment was performed at Stn. ALOHA (22°45' N 158° W) in the NPSG during KM1016 cruise (R/V Kilo-Moana). Surface water was collected on August 22, 2010 from 10 m depth using Niskin bottles in a rosette mounted to a conductivity-temperature-depth (CTD) instrument (cast S2c9), filtered through 10.0 μm pore-size mesh, and distributed into twelve 4 L clear polycarbonate bottles. The bottles and tubing were cleaned with10% HCl, but vigorous trace-metal clean precautions were

not taken during experimental setup. Each treatment was done in triplicate: 1) control (no enrichment), 2) enrichment with 1.0 μM K_2PO_4, and 3) enrichment with 2.0 nM $FeCl_3$. The bottles were incubated in a deck incubator continuously flushed with surface seawater to maintain the proper temperature. Neutral density screening was used to attenuate sunlight to ca. 35% of surface sunlight. Four L samples were taken from the original seawater sample before nutrient additions (three replicates total) and from all treatments after 48 h of incubation in the morning. From each bottle, 3.9 L were then filtered onto Sterivex cartridges (0.22 μm, Millipore, Billerica, MA) using gentle Masterflex (Cole Parmer, Vernon Hills, IL) peristaltic pumping ensuring that filtration time did not exceed 25 min. Sterivex cartridges were immediately flash frozen in liquid nitrogen and stored at -80°C until processing. The remaining 0.1 L from each bottle was used in Fast Repetition Rate Fluorometer (FRRF) analysis.

RNA extraction and processing for hybridization to the microarray.

RNA was extracted using the Ambion® RiboPure™ kit (Life Technologies, Grand Island, NY) with modifications that included mechanical lysis using glass beads (Supplementary Material). Extracted RNA was treated with DNase to remove genomic DNA. RNA quantity and quality were determined with a NanoDrop 1000 (Thermo Scientific, Waltham, MA) and a 2100 Bioanalyzer (Agilent Technologies, Santa Clara, CA, USA) using the RNA 6000 Nano kit (Agilent Technologies). Only samples with RNA Integrity Number (RIN) >7.0 and ratios of A260/A230 and A260/A280 ≥1.8 were processed further. RNA yield from the incubations of the surface community at Stn. ALOHA ranged from 280 to 1130 ng L^{-1}., an amount insufficient for the NimbleGen microarray. cDNA was synthesized from 400 ng RNA from each sample, which was

then amplified using the TransPlex Whole Transcriptome Amplification kit (WTA-1, Sigma-Aldrich, St Louis, MO, USA) and antibody inactivated hot-start Taq DNA Polymerase (Sigma-Aldrich). The amplified cDNA was purified with the GenElute PCR cleanup kit (Sigma-Aldrich), and the quality and quantity of double-stranded (ds) cDNA was determined with NanoDrop 1000 and a 2100 Bioanalyzer using the Agilent DNA 7500 kit (Agilent Technologies). Four hundred ng of total RNA yielded on average 12 μg of ds-cDNA. The amplification efficiency was determined with a TaqMan qPCR assay targeting a spike-in transcript ERCC-00116 (Invitrogen, Life Technologies). One μL of 1:100 dilution (corresponding to 4.7 attomoles of ERCC-0016) of the ERCC (External RNA Control Consortium, Lemire *et al*., 2011) RNA spike-in mix 1 (Ambion®) was added to RNA samples before amplification. Amplification of one replicate for Fe treatment was seven times less than the average amplification (150-fold), and the sample was excluded from microarray hybridization. The labeling and hybridization of cDNA samples (1.0 μg of ds-cDNA) to the microarray was done at the Sandler Center Functional Genomics Core Facility (University of California, San Francisco, CA, USA) according to the protocol in NimbleGen Arrays User's Guide: "Gene Expression Arrays, version 6.0".

DNA extraction and estimation of cell abundances for diazotrophs and Synechococcus spp.

DNA was extracted from the organic phase of the nucleic acid extract after RNA separation using the RiboPure™ kit (Ambion) according to the manufacturer's instructions (Supplementary Material). DNA yield ranged from 9.2 to 26.6 μg L^{-1}. Diazotroph abundances were determined with Taqman® quantitative PCR (qPCR) assays

targeting *nifH* as described in Moisander *et al.* (2010). *Synechococcus* spp. cell abundances were estimated using the qPCR assay targeting *narB* (group G) as described in Paerl *et al.* (2012). The Tukey's test was used to compare the qPCR data.

Microarray data analysis

All data analyses were performed with R (www.R-project.org) and the Bioconductor project (Gentleman *et al.*, 2004), specifically using the limma (Smyth, 2005), affy (Gautier *et al.*, 2004), and samr (Tusher *et al.*, 2001) packages; plots were made using gplots package. Transcription values were obtained using the robust multi-array average (RMA) algorithm (Irizarry *et al.*, 2003) and using Li-Wong across-chip normalization (Li and Wong, 2001) (Figure S2). The signal to noise ratio (SNR) of each chip was calculated as: $SNR=(S_i - BG)/BG$; where S_i – hybridization signal for the gene, BG – chip background signal determined as average of the lowest 5% of all signals. Transcription was considered detected if SNR of a transcript was ≥ 5 The detection range was estimated based on the ERCC hybridization data (Figure S3). The detection limit as percent of total mRNA was calculated for 1000 nucleotide long mRNA and considering that rRNA constitutes 95% of total RNA. The relative cell sensitivity limit was estimated based on the assumption of 1380 mRNA per cell (Neidhardt, 1996). For each group (Eukaryota, Bacteria non-picocyanobacteria, *Prochlorococcus*, *Synechococcus*, Viruses, Archaea), gene transcription was scaled to the median of the group in each sample. To identify differentially transcribed genes, the nonparametric method, Significance Analysis of Microarray (SAM, Tusher *et al.*, 2001) was used with the following parameter settings: delta=0.3, 100 iterations, False Discovery Rate (FDR)=0.05 (Figure S4). In addition, Linear Models for Microarray (LIMMA, Smyth, 2005) was used with

the following parameters: fold change in $\log_2$ scale (FC) =1; FDR=0.05; $p<0.1$ (Benjamini-Hochberg adjusted, Benjamini and Hochberg, 1995). The separation of samples by treatment was supported by a Wilcoxon test (Bauer, 1972) performed with at least a hundred random resamplings of 1,000 gene probes (Figure S4). Transcription data was centered and scaled across genes, and a distance matrix was calculated by Pearson's correlation coefficient. The distance matrix was then used in hierarchical clustering by a complete agglomeration method. Raw and normalized microarray data were prepared in accordance to MIAME standards (Brazma *et al.*, 2001) and submitted to NCBI GEO under accession number GSE44448.

Fast Repetition Rate Fluorometer (FRRF) *measurements*

Chlorophyll *a* variable fluorescence (Fv) and maximal fluorescence (Fm) were measured using FRRF as described in Kolber *et al.* (1998). FRRF measurements were taken for each sample in the beginning and after 48 h of incubation in six replications and using blue light (470 nm) for excitation. The Tukey's test was used to compare the FRRF data.

Results and Discussion

Microarray design

The MicroTOOLs microarray targets marker genes for three domains of life in marine microbial communities along with their known viruses. Using our approach, a total of 19,400 target sequences representing 145 genes (Table 1) were obtained from genomes (~8%) and from metagenomes/metatranscriptomes (~92%). The design was biased towards picocyanobacteria *Prochlorococcus* and *Synechococcus* spp., and the Alphaproteobacteria clade SAR11 sequences (Figure 1, Table S1) due to their high

abundances in surface oceanic waters, resulting in high abundances of their nucleotide sequences in metagenomes and metatranscriptomes (e.g. DeLong *et al*., 2006; Hewson *et al.,* 2010) and the fact that gene function and diversity of these microorganisms have been relatively well-studied (Scanlan and West, 2002; Scanlan, 2009; Sun *et al.*, 2011; Brown *et al*., 2012). Genes from less abundant prokaryotic microorganisms, such as from marine N_2-fixing cyanobacteria, were also included. Marine eukaryotic phytoplankton were primarily represented by the genes encoding the large subunit of RuBisCO (*rbcL*) and nitrate reductase (NR) (Figure 1). Additional eukaryotic genes were selected based on available genomes and/or expressed sequence tags (EST) for diatoms (Armbrust et al 2004; Bowler et al 2008). Genes for marine DNA and RNA viruses included genes for DNA polymerase, major capsid protein, and RNA-dependent RNA polymerase. Probes for Archaea targeted genes encoding ammonia monooxygenase, RuBisCO, and urease (Table S1).

As a proof of concept, the transcriptional responses of the surface microbial community at Stn. ALOHA to phosphate (P) or iron (Fe) amendments were analyzed using the MicroTOOLs microarray. During the time of sampling, surface waters at Stn. ALOHA had 54-79 nmol L^{-1} of P, 4-33 nmol L^{-1} of nitrate plus nitrite, 0.79-1.07 μmol L^{-1} of silicon, and 68-78 ng L^{-1} of chlorophyll *a* (data from 5-10m depths). Fe concentrations were not measured during the time of study, but average near surface Fe concentration at Stn. ALOHA is 0.44 nmol L^{-1} (Boyle et al 2005). The long-term mean P concentrations at Stn. ALOHA are 58±3 nmol L^{-1} (Björkman *et al.*, 2012), and thus, out of P and Fe, Fe was believed to be the limiting nutrient at the time of incubation, at least for diazotrophic community.

Microbial community transcription at Stn. ALOHA detected with the MicroTOOLs microarray

Hybridization signals above background were detected for a total of 15,507 genes (68% of the microarray set) with an average detection of 40% of target orthologs for each gene (Table S3). The range of detection for the microarray was from 700 to 11E+06 transcript copies based on the spike-in ERCC (External RNA Control Consortium) data (Figure S3). Calculated for a 1000 nucleotide long mRNA, the 700 transcript copies (absolute sensitivity) corresponded to 1.8E-06 % of the total community mRNA and to 0.0025% as the lowest relative abundance of cells within the community that can be detected. Average transcription from pelagophytes (*Aureococcus*) and prymnesiophytes (*Chrysochromulina, Phaeocystis, Helicosphaera*) was up to 6-fold higher than the median transcription in all eukaryotes across samples (Figure 2A, Table S2). Average transcription among prokaryotes (*Prochlorococcus, Synechococcus,* Proteobacteria,) was distributed around the median transcription of the corresponding transcriptome in all samples (Figure 2A, insert). The exceptions were genes related to an uncultured *Prochlorococcus* species represented by a fosmid clone HOT0M, which had 16-fold higher transcription than the median (Figure 2A). The most highly transcribed genes across all treatments were *psaA* (photosystem I)*, psbA* (photosystem II), *amt* (ammonium transport), *urtA* (urea transporter), and *rbcL* (Figure 2B). Among relatively low abundance taxa, diazotrophic cyanobacteria, *Trichodesmium erythraeum* IMS101, *Candidatus* Atelocyanobacterium thalassa (*Ca*. A. thalassa, or unicellular cyanobacterial group A, UCYN-A)*,* and an uncultured heterocystous cyanobacterium (NCBI 112280460, Moisander *et al.*, 2007) had high *nifH* (Fe-nitrogenase reductase)

transcription (Table S2). The maximum *nifH* transcription in these cyanobacteria occurs during the early morning hours (Church *et al.*, 2005), when samples from incubations were collected. Another N_2-fixing cyanobacterium *Crocosphaera watsonii (Crocosphaera)* had high transcription of *pstS* (high-affinity phosphate-binding) and *ftsZ* (cell division) (Table S2). *Ca*. A. thalassa and *Crocosphaera* cell abundances (Table 2) in the incubations ranged from 0.0002% to 0.04% of the total prokaryotic community (8.0E+08 cells L^{-1}, Bjorkmann *et al.,* 2012) overlapping the estimated relative cell detection limit for the microarray (0.0025%). In addition, gene transcripts were detected for members of the *Phycodnavirus* family, which infects a number of eukaryotic phytoplankton including *Micromonas* (Mayer and Taylor, 1979) and *Aureococcus* (Milligan and Cosper, 1994), and also for cyanophages from the *Myoviridae* family (DNA polymerase and viral capsid genes).

Overall, the genes with detected transcription reflected the composition and activity of the microbial community at Stn. ALOHA previously described (DeLong *et al.,* 2006; Frias-Lopez *et al.,* 2008; Hewson *et al.*, 2010; Church et al., 2009). Moreover, the high activity of the pelagophytes and prymnesiophytes was consistent with the detection of a eukaryotic phytoplankton bloom at that time (Björkman *et al.*, 2012).

Transcription by members of unknown lineages

A wide diversity of genotypes of *Prochlorococcus*, *Synechococcus*, and Alphaproteobacteria were detected at the study site. *Prochlorococcus* probes that yielded detectable signals had a wide range of nucleotide similarity to sequenced genomes (Figure 3A, D) indicating a broad representation of the known natural genetic diversity in *Prochlorococcus*, as well as transcriptional activity in genotypes with no currently

sequenced genome. Gene probes with detectable transcription had a median of 91% nucleotide identity to *Prochlorococcus* genome sequences, such as strain CCMP1986 (Figure 3A).

Detected *Synechococcus*-like transcripts had a median of 88.5% similarity at the nucleotide level to their orthologs in known genomes (Figure 3B, D). Especially high transcriptional activity (normalized transcription >2) was detected for genes similar to *Synechococcus* sp. RCC307, a strain from *Synechococcus* sub-cluster 5.3A. The low percent similarity to known genomes indicates the existence of an uncharacterized lineage, potentially within the sub-cluster 5.3, at Stn. ALOHA. The presence of this clade has recently been reported in the open ocean and in the Mediterranean Sea, with abundances correlated to warm, low-nutrient waters (Ahlgren and Rocap, 2012; Huang *et al.*, 2012; Mella-Flores *et al.*, 2011; Post *et al.*, 2011).

With a median of 80.0% similarity, probe sequences targeting Alphaproteobacteria also displayed a degree of degeneracy relative to known genome sequences (Figure 3C), suggesting that new, active strains remain uncharacterized. This is consistent with another recent study, where a new group of the SAR11 clade was proposed based on 16S rRNA gene phylogeny (Allen *et al.*, 2012). Such uncharacterized genotypes with detectable transcription are candidates for further targeted genomic studies.

Response to nutrient amendments.

Maximum chlorophyll *a* fluorescence (Fm) was not significantly different between each of the amendments and the control and between P and Fe amendments ($p>0.2$, Table 2). The lower ratio of variable to maximum fluorescence (Fv/Fm) in the Fe

treatment ($p<0.05$) may be a result of either a shift in phytoplankton community composition or lower photosynthetic efficiency in Fe-enriched samples due to nutrient limitation (Sylvan *et al.*, 2007; Sylvan *et al.*, 2011; Vogel *et al.*, 2003).

Cyanobacterial abundances increased in response to Fe amendments. Diazotrophic cyanobacteria *Crocosphaera* and *Ca.* A. thalassa and the non-diazotrophic cyanobacteria *Synechococcus* were 8, 19, and 23 times, respectively, more abundant in Fe-amendments versus the control after incubation as measured with qPCR (Table 2). *Ca.* A. thalassa and *Synechococcus* spp. were 8 and 11 times, respectively, more abundant in Fe-amended versus P-amended samples. This high increase in cell abundances in response to Fe and low response to P is consistent with previously reported variability in responses of diazotrophs to P and Fe availability in the NPSG (Zehr *et al.*, 2007; Grabowski et al., 2008; Watkins-Brandt et al., 2011).

Differential transcriptional responses to P and Fe amendments

After 48 h of incubation, Fe and P amendments yielded significant differences in transcript levels across the microbial community despite the high variability among biological replicates. When transcription of all genes was compared, biological replicates had weak positive to weak negative correlation (Table S4). Strong correlations between replicates were obtained at the level of specific phylogroups (eukaryotic, *Prochlorococcus*, and *Synechococcus*) and metabolic functions (energy and N metabolism) (Figure 4). Factors that may have caused discrepancies in transcription profiles between replicates include differences in biological processes within individual bottles (such as protist grazing, viral lysis) and biases of sample collection and processing.

The SAM analysis (Significance Analysis of Microarray, see Methods) identified 3,742 genes as significantly differentially transcribed between treatments (FDR=0.05) from a total of 15,507 genes with detected transcription (Figure S5, Table S5). Overall, the addition of Fe resulted in the increased transcription of 1699 genes encoding for N metabolism, photosynthesis, oxidative phosphorylation and ABC-type transporters (Figure 4F). The category of genes not assigned to a KEGG pathway (NA) and up-regulated in the Fe-amendment included genes for ammonium transport, organic P assimilation, DNA replication, and cell division. The 2043 gene up-regulated in the P amendment were enriched in genes for carbon fixation (Figure 4G).

It is important to note that highly transcribed genes in all samples (for example, eukaryotic nitrate reductase NCBI GI: GU203403 (Table S2) oversaturated the hybridization signal, and the difference in transcription could not be estimated.

Relief from Fe limitation in oligotrophic taxa

Transcriptional patterns showed that Fe amendment resulted in relief from Fe limitation in taxa common to oligotrophic waters.

Transcription of genes for energy metabolism increased in picocyanobacteria upon Fe amendment (Figure 5A and B). Fe additions resulted in two-fold increased transcription of *psaA* (photosystem I) in both *Prochlorococcus* and *Synechococcus* spp, with an up to five-fold increase in transcript level for the *psaA* gene in the *Synechococcus* RCC307-like genotype. Transcription of the *coxA* gene (cytochrome C oxidase) in *Prochlorococcus* eHLI and eHLII genotypes was up to five-fold higher in the Fe-amended treatment. The role of Fe as a cofactor in electron transport is consistent with increases in transcripts for energy metabolism genes in *Prochlorococcus* and

Synechococcus in Fe amendments, a pattern also observed in cultures (e.g. Singh *et al.*, 2003; Thomposon *et al.*, 2011).

This increased transcription of energy metabolism genes was accompanied by the increased transcription of genes for cellular activity (DNA replication *recA*, cell division *ftsZ*) in *Prochlorococcus* (high light ecotype I, eHLI) in Fe amended treatments (Figure S6). It is possible that addition of Fe resulted in the increased growth rate of *Prochlorococcus* as observed in the Eastern South Pacific (Mann and Chisholm, 2000). The higher sensitivity of the HLI ecotype in comparison to LL ecotypes to Fe availability (Thompson *et al.*, 2011) may explain the increased cellular activity of *Prochlorococcus* eHLI in response to the Fe amendment. The availability of fixed N from diazotrophs (Mullholland *et al*., 2004; Mullholland and Capone, 2001) upon Fe-amendment may have been an additional or alternative factor that enhanced the growth of *Prochlorococcus* eHLI in these treatments. This hypothesis is consistent with the similarity in responses of diazotrophs and *Prochlorococcus* HL spp. to Fe addition, both in this study and in a study in the Southwest Pacific Ocean (Moisander *et al.*, 2012).

The down-regulation of *idiA* in oligotrophic taxa such as *Prochlorococcus* AS9601 (eHLII) and Alphaproteobacteria SAR11 clade strain HIMB59 (Figure S6) in Fe-amendments was consistent with the repression of the gene in Fe-replete cultures (Bagg and Neilands, 1987; Smith *et al.*, 2010; Thompson *et al.*, 2011; Webb *et al.*, 2001). In contrast, *idiA* transcription by more eutrophic *Synechococcus* RCC307-like (sub-cluster 5.3A, Mella-Flores *et al.*, 2011) and *Pelagibacter* HTCC7211-like genotypes was higher in the Fe-amendment. The differences between coastal and open ocean species have been reported before in cultures and include differences in Fe requirements and

sensing (e.g. Palenik *et al.*, 2006; Sunda *et al.*, 1991), uptake of siderophore- or porphyrin-bound Fe (Hutchins *et al.*, 1999), and post-translational regulation by antisense RNA (Hernández *et al.*, 2006).

This is the first report that validates differential Fe responses across a multitude of taxa in a mixed community. The relatively high (0.44 nmol L^{-1}) average Fe concentrations in the surface waters at Stn. ALOHA are not usually considered limiting for microbial communities (Boyle *et al.,* 2005). The response to Fe amendments reported here indicates that either Fe concentrations at the time of study were lower than average or that much of this Fe was not bioavailable, at least for some taxa.

Increased N metabolism and N limitation in Fe amendments

The availability of Fe led to an increased transcription of N transport and metabolism genes (*amt*, *urtA*, *nrtP*, *glnA*, *nirA*, *sigAII*) in both *Synechococcus* and *Prochlorococcus* and possibly led to N limitation in *Prochlorococcus* eHL (Figure 5C, Table S6). The observed four-fold down-regulation of the *rbcL* transcripts in *Prochlorococcus* eHLI and eHLII in the Fe amendment (Figure 5D) might have resulted from N limitation (Tolonen *et al.*, 2006). In contrast, *Synechococcus* spp. had two-fold higher *rbcL* transcription in Fe-amendments than in P-amendments (Table S6). We speculate that the differences in *rbcL* transcription were due to the ability of the majority of *Synechococcus* spp. to assimilate nitrate, while only few uncultured *Prochlorococcus* spp. have that capability (Moore *et al*., 2002; Martiny *et al.*, 2009). Transcription of nitrite/nitrate utilization genes (*nrtP*, *nirA*, and *narB*) was detected in *Synechococcus* spp. and was up-regulated in *Synechococcus* WH8109 (clade II) in Fe amendments (Table S2 and S6). Transcription of the *narB* gene was also detected for a few uncultured

Prochlorococcus spp., but the precise affiliation of these genes is unknown. While not measured directly in the incubations, nitrate/nitrite concentrations at Stn. ALOHA during incubation were relatively high. It is possible that *Synechococcus* spp. were able to obtain sufficient nitrogen, and thus carbon fixation in *Synechococcus* was not down-regulated.

The eukaryotic nitrate reductase gene (NR) was up-regulated in Fe-amendments (Table S6) possibly resulting from increased energy production, by sufficient Fe requirements for the nitrate reductase enzyme, and/or by subsequent N limitation. Although transcription of NR can be induced by nitrate alone and not by N limitation in eukaryotes (Poulsen and Kröger, 2005; Song and Ward, 2004), the lower Fv/Fm ratio was possibly due to insufficient N (Tolonen *et al.*, 2006) for dominant phytoplankton in chlorophyll *a* measurements, which at that time were eukaryotic phytoplankton. Similar to *Prochlorococcus*, eukaryotic phytoplankton down-regulated *rbcL* genes in the Fe treatment, especially two Chrysophytes, *Epipyxis pulchra* and *Ochromonas aestuarti*, and two Prymnesiophytes, *Chrysochromulina alifera* and *Chrysochromulina flava* (Figure 5D). The down-regulation of *rbcL* could be due to reduced cellular N:C ratio in these organisms. Alternatively, if photosynthetic eukaryotic phytoplankton were utilizing nitrate, they would have had to divert ATP and reductant away from carbon fixation, which would also result in decreased carbon fixation (Laws, 1991) by down-regulating *rbcL* transcription.

P limitation or increased P metabolism followed Fe addition

Individual taxa in the Fe treatment showed up-regulated transcription of P stress response genes, despite the generally high availability of phosphate. Transcription of *pstS* (the high-affinity phosphate binding protein) in *Prochlorococcus* MIT9301-like

genotypes (eHLII) was two-fold higher in the Fe-amendments in comparison to the P-amendments (Figure S6). The set of iron up-regulated genes was enriched in *pstS* transcripts from Proteobacteria (especially, Alphaproteobacterium HIMB5) and cyanobacteria (*Synechococcus* WH8102) (Table S6). The phosphonate utilization genes *phnJ* and *phnA* in Alphaproteobacteria and the alkaline phosphatase gene *phoD* in Gammaproteobacteria were up-regulated in the Fe addition (Figure S6, Table S6). The 2 m size fraction of phytoplankton was not P-depleted at the time of incubation (Björkman *et al.*, 2012), but it is possible that P limitation was induced by the fast removal of P in the Fe amendment by some members of the community such as nitrate utilizers. Alternatively, increased energy generation in microorganisms in the Fe treatment might have provided energy for phosphate acquisition and membrane translocation systems for proteins involved in alternative P source assimilation (Cembella *et al.*, 1982; Jansson, 1988; Tetu *et al.*, 2009). Taxa that up-regulated P transport and metabolism genes were largely different from taxa that demonstrated N limitation, but manifestations of both N and P limitations occurred after Fe was supplied suggesting Fe as a primary limiting nutrient at that time.

Conclusions

This study reports the design of a high-density oligonucleotide microarray (MicroTOOLs) that targets marine microbial communities which was informed by currently available environmental sequence data. We detected differential microbial community responses to nutrient amendments in the NPSG, ultimately demonstrating strain-specific community responses to relief of Fe stress that was followed by N or P

limitation in some taxa. Such data provide a mechanistic understanding of changes in microbial communities in response to nutrient fluxes or other environmental factors.

Future technological improvements, such as automated probe design, including probes for other marker genes and exploratory probes (Chung *et al.*, 2005; Dugat-Bony *et al.*, 2011) would improve the performance of the microarray. In addition to the utility of the MicroTOOLs array for incubation experiments, this microarray could be applied as a tool for pelagic marine microbiological studies for standardized information across study types and ocean basins. This would result in a high-resolution map of microbial genes and their transcriptional activities in the environment, and provide the baseline for assessing the impacts of future perturbations of the global ocean.

Acknowledgments

This work resulted from two workshops and the design of the microarray was partially supported by the Gordon and Betty Moore Foundation (the MEGAMER facility), by a Gordon and Betty Moore Foundation Marine Investigator grant (JPZ) and by the NSF Center for Microbial Oceanography (C-MORE, NSF EF0424599). We thank all participants of the October 2010 MicroTOOLs workshop sponsored by the MOORE foundation. Particularly, we thank those who provided sequences for the design of microarray: Mahdi Belcaid, Dreux Chappell, Jackie Collier, Chris Francis, Scott Gifford, Jana Grote, Bethany Jenkins, Julie LaRoche, Pia Moisander, Annika Mosier, Micaela Parker, Holly Simon, Mariya Smit, Jody Wright, and Louie Wurch. We thank Shulei Sun and the team at CAMERA for support during microarray design. We also thank the Hawai'i Ocean Time-series (HOT) program and C-MORE, specifically the captain and

crew of R/V Kilo Moana and chief scientist of KM01016 cruise, Matt Church, for their expertise, and for providing the opportunity and support for conducting experiments at sea. We thank Philip Heller and Jonathan Magasin for developing Java scripts for the microarray design and for computational support during the MicroTOOLs workshop in October 2010. Finally, we thank Nicole Pereira for help running the incubation experiment, Sasha Tozzi for FRRF measurements, Rob Tibshirani for a consultation on microarray data normalization and analysis, and the JPZ laboratory members for constructive discussions.

Conflicts of interest

Authors declare no conflict of interest in relation to the submitted work.

Supplementary information is available at the ISME Journal's website.

References

Abell GCJ, Robert SS, Frampton DMF, Volkman JK, Rizwi F *et al.* (2012) High-throughput analysis of ammonia oxidiser community composition via a novel, *amoA*-based functional gene array. *PLoS ONE* **7**: e51542. E51542.

Ahlgren NA, Rocap G. (2012). Diversity and distribution of marine *Synechococcus*: multiple gene phylogenies for consensus classification and development of qPCR assays for sensitive measurement of clades in the ocean. *Front Microbio* **3:** 213.

Allen LZ, Allen LZ, Badger JH, McCrow JP, Paulsen IT, Elbourne LDH *et al.* (2012). Influence of nutrients and currents on the genomic composition of microbes across an upwelling mosaic. *ISME J* **6**:1403-1414.

Armbrust EV, Berges JA, Bowler C, Green BR, Martinez D, Putnam NH *et al.* (2004). The genome of the diatom *Thalassiosira Pseudonana*: Ecology, Evolution, and Metabolism. *Science* **306**: 79-86.

Azam F, Malfatti F. (2007). Microbial structuring of marine ecosystems. *Nat Rev Microbiol* **5:** 782-791.

Bagg A, Neilands JB. (1987). Ferric uptake regulation protein acts as a repressor, employing iron (II) as a cofactor to bind the operator of an iron transport operon in *Escherichia coli*. *Biochemistry* **26:** 5471-5477.

Bai S, Li J, He Z, Van Nostrand JD, Tian Y, Lin G *et al.* (2013). GeoChip-based analysis of the functional gene diversity and metabolic potential of soil microbial communities of mangroves. *Appl Microbiol Biotechnol* 97: 7035-7048.

Bauer DF. (1972). Constructing confidence sets using rank statistics. *J Am Stat Assoc* **67:** 687-690.

Benjamini Y, Hochberg Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J R Stat Soc Ser B (Methodological)* **57:** 289-300.

Björkman K, Duhamel S, Karl DM. (2012). Microbial group specific uptake kinetics of inorganic phosphate and adenosine-5'-triphosphate (ATP) in the North Pacific Subtropical Gyre. *Front Microbiol* **3**: 189.

Bouskill NJ, Eveillard D, O'Mullan G, Jackson GA, Ward BB. (2011). Seasonal and annual reoccurrence in betaproteobacterial ammonia-oxidizing bacterial population structure. *Environ Microbiol* **13:** 872-86.

Bowler C, Allen AE, Badger JH, Grimwood J, Jabbari K, Kuo A *et al.* (2008). The *Phaeodactylum* genome reveals the evolutionary history of diatom genomes. *Nature* **456**: 239–244.

Boyle EA, Bergquist BA, Kayser RA, Mahowald N. (2005). Iron, manganese, and lead at Hawaii Ocean Time-series station ALOHA: Temporal variability and an intermediate water hydrothermal plume. *Geochim Cosmochim Acta* **69**: 933-952.

Brazma A, Hingamp P, Quackenbush J, Sherlock G, Spellman P, Stoeckert C *et al.* (2001). Minimum information about a microarray experiment (MIAME) - toward standards for microarray data. *Nat Genet* **29:** 365-371.

Brodie EL, DeSantis TZ, Joyner DC, Baek SM, Larsen JT, Andersen GL *et al.* (2006). Application of a high-density oligonucleotide microarray approach to study bacterial population dynamics during uranium reduction and reoxidation. *Appl Environ Microbiol* **72:** 6288-6298.

Brown MV, Lauro FM, DeMaere MZ, Muir L, Wilkins D, Thomas T *et al.* (2012) Global biogeography of SAR11 marine bacteria. *Mol Syst Biol* **8:** 595.

Bulow SE, Francis CA, Jackson GA, Ward BB. (2008). Sediment denitrifier community composition and *nirS* gene expression investigated with functional gene microarrays. *Environ Microbiol* **10:** 3057-3069.

Carpenter EJ, Romans K. (1991). Major role of the cyanobacterium *Trichodesmium* in nutrient cycling in the North Atlantic Ocean. *Science* **254:** 1356-1358.

Cembella AD, Antia NJ, Harrison PJ. (1982). The utilization of inorganic and organic phosphorous compounds as nutrients by eukaryotic microalgae: A multidisciplinary perspective: Part I. *Crit Rev Microbiol* **10:** 317-391.

Chen F, Wang K, Kan JJ, Bachoon DS, Lu JR, Lau S *et al.* (2004). Phylogenetic diversity of *Synechococcus* in the Chesapeake Bay revealed by ribulose-1,5-bisphosphate carboxylase-oxygenase (RuBisCO) large subunit gene (*rbcL*) sequences. *Aquat Microb Ecol* **36:** 153-164.

Chung WH, Rhee SK, Wan XF, Bae JW, Quan ZX, Park YH. (2005). Design of long oligonucleotide probes for functional gene detection in a microbial community. *Bioinformatics* **21**, 4092–4100.

Church MJ, Mahaffey C, Letelier RM, Lukas R, Zehr JP, Karl DM. (2009). Physical forcing of nitrogen fixation and diazotroph community structure in the North Pacific Subtropical Gyre. *Glob Biogeochem Cycles* **23**: GB2020.

Church MJ, Short CM, Jenkins BD, Karl DM, Zehr JP. (2005). Temporal patterns of nitrogenase gene (*nifH*) expression in the oligotrophic North Pacific Ocean. *Appl Environ Microbiol* **71:** 5362-5370.

Coleman ML, Chisholm SW. (2010). Ecosystem-specific selection pressures revealed through comparative population genomics. *Proc Natl Acad Sci USA* **107:** 18634-18639.

DeLong EF, Preston CM, Mincer T, Rich V, Hallam SJ, Frigaard NU *et al*. (2006). Community genomics among stratified microbial assemblages in the ocean's interior. *Science* **311:** 496-503.

Dugat-Bony E, Peyretaillade E, Parisot N, Biderre-Petit C, Jaziri F, Hill D *et al.* (2011). Detecting unknown sequences with DNA microarrays: explorative probe design strategies. *Enviro. Microbiol* **14:** 356-371.

Dupont CL, Barbeau K, Palenik B. (2008). Ni uptake and limitation in marine *Synechococcus* strains. *Appl Environ Micro* **74**: 23-31.

Dyhrman ST, Haley ST. (2006). Phosphorus scavenging in the unicellular marine diazotroph *Crocosphaera watsonii*. *Appl Environ Microbiol* **72:** 1452-1458.

Feingersch R, Suzuki MT, Shmoish M, Sharon I, Sabehi G, Partensky F *et al.* (2010). Microbial community genomics in eastern Mediterranean Sea surface waters. *ISME J* **4:** 78-87.

Frias-Lopez J, Shi Y, Tyson GW, Coleman ML, Schuster SC, Chisholm SW *et al.* (2008). Microbial community gene expression in ocean surface waters. *Proc Natl Acad Sci USA* **105:** 3805-3810.

Fuhrman JA, Hewson I, Schwalbach MS, Steele JA, Brown MV, Naeem S. (2006). Annually reoccurring bacterial communities are predictable from ocean conditions. *Proc Natl Acad Sci USA* **103:** 13104-13109.

Fuller NJ, West NJ, Marie D, Yallop M, Rivlin T, Post AF *et al*. (2005). Dynamics of community structure and phosphate status of picocyanobacterial populations in the Gulf of Aqaba, Red Sea. *Limnol Oceanogr* **50:** 363-375.

Gautier L, Cope L, Bolstad BM, Irizarry RA. (2004). affy - analysis of Affymetrix GeneChip data at the probe level. *Bioinformatics* **20:** 307-315.

Gentleman RC, Carey VJ, Bates DM, Bolstad B, Dettling M, Dudoit S *et al.* (2004). Bioconductor: Open software development for computational biology and bioinformatics. *Genome Biol* **5:** R80.

Giovannoni SJ, Hayakawa DH, Tripp HJ, Stingl U, Givan SA, Cho J-C *et al.* (2008). The small genome of an abundant coastal ocean methylotroph. *Environ Microbiol* **10:** 1771-1782.

Giovannoni SJ, Vergin KL. (2012). Seasonality in ocean microbial communities. *Science* **335:** 671-676.

Grabowski M, Church MJ, Karl DM. (2008). Nitrogen fixation rates and controls at Stn ALOHA. *Aquat Microb Ecol* **52**: 175–183.

Halsey KH, Carter AE, Giovannoni SJ. (2012). Synergistic metabolism of a broad range of C1 compounds in the marine methylotrophic bacterium HTCC2181. *Environ Microbiol* **14:** 630-640.

He Z, Deng Y, Van Nostrand JD, Tu Q, Xu M, Hemme CL *et al.* (2010). GeoChip 3.0 as a high-throughput tool for analyzing microbial community composition, structure and functional activity. *ISME J* **4:** 1167-1179.

He Z, Gentry TJ, Schadt CW, Wu L, Liebich J, Chong SC *et al.* (2007). GeoChip: a comprehensive microarray for investigating biogeochemical, ecological and environmental processes. *ISME J* **1:** 67-77.

Hernández JA, Muro-Pastor AM, Flores E, Bes MT, Peleato ML, Fillat MF. (2006). Identification of a *furA cis* antisense RNA in the cyanobacterium *Anabaena* sp. PCC 7120. *J Mol Biol* **355:** 325-334.

Hewson I, Poretsky RS, Tripp HJ, Montoya JP, Zehr JP. (2010). Spatial patterns and light-driven variation of microbial population gene expression in surface waters of the oligotrophic open ocean. *Environ Microbiol* **12:** 1940-1956.

Holtzendorff J, Marie D, Post AF, Partensky F, Rivlin A, Hess WR. (2002). Synchronized expression of *ftsZ* in natural *Prochlorococcus* populations of the Red Sea. *Environ Microbiol* **4:** 644-653.

Huang S, Wilhelm SW, Harvey HR, Taylor K, Jiao N, Chen F. (2012). Novel lineages of *Prochlorococcus* and *Synechococcus* in the global oceans. *ISME J* **6:** 285-297.

Huang Y, Niu B, Gao Y, Fu L, Li W. (2010). CD-HIT Suite: a web server for clustering and comparing biological sequences. *Bioinformatics* **26:** 680-682.

Hutchins DA, Witter AE, Butler A, Luther GW. (1999). Competition among marine phytoplankton for different chelated iron species. *Nature* **400:** 858-861.

Ilikchyan IN, McKay RM, Zehr JP, Dyhrman ST, Bullerjahn GS. (2009). Detection and expression of the phosphonate transporter gene *phnD* in marine and freshwater picocyanobacteria. *Environ Microbiol* **11**: 1314-1324.

Irizarry RA, Hobbs B, Collin F, Beazer-Barclay YD, Antonellis KJ, Scherf U *et al.* (2003). Exploration, normalization, and summaries of high density oligonucleotide array probe level data. *Biostatistics* **4:** 249-264.

Jansson M. (1988). Phosphate uptake and utilization by bacteria and algae. *Hydrobiologia* **170:** 177-189.

Kamennaya NA, Post AF. (2011). Characterization of cyanate metabolism in marine *Synechococcus* and *Prochlorococcus* spp. *Appl Environ Microbiol* **77:** 291-301.

Karl DM, Letelier R. (2008). Nitrogen fixation-enhanced carbon sequestration in low nitrate, low chlorophyll seascapes. *Mar Ecol Prog Ser* **364:** 257-268.

Karl DM, Letelier R, Tupas L, Dore J, Christian J, Hebel D. (1997). The role of nitrogen fixation in biogeochemical cycling in the subtropical North Pacific Ocean. *Nature* **388**: 533–538.

Karl DM, Lukas R. (1996). The Hawaii ocean time-series (HOT) program: Background, rationale, and field implementation. *Deep-Sea Research II* **43:** 129-156.

Kolber ZS, Prasil O, Falkowski PG. (1998). Measurements of variable chlorophyll fluorescence using fast repetition rate techniques: defining methodology and experimental protocols. *Biochim Biophys Acta-Bioenerg* **1367:** 88-106.

Laws EA. (1991). Photosynthetic quotients, new production and net community production in the open ocean. *Deep-Sea Res Oceanogr Part A* **38:** 143-167.

Lemire A, Lea K, Batten D, Gu J, Whitley P, Bramlett K. (2011). Development of ERCC RNA spike-In control mixes. *J Biomol Tech* **22**: S46.

Li C, Wong WH. (2001). Model-based analysis of oligonucleotide arrays: expression index computation and outlier detection. *Proc Natl Acad Sci USA* **98:** 31-36.

Li W, Godzik A. (2006). CD-HIT: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics* **22:** 1658 - 1659.

Lindell D, Post AF. (2001). Ecological aspects of *ntcA* gene expression and its use as an indicator of the nitrogen status of marine *Synechococcus* spp. *Appl Environ Microbiol* **67:** 3340-3349.

Lu Z, Deng Y, Van Nostrand JD, He Z, Voordeckers J, Zhou A *et al.* (2012). Microbial gene functions enriched in the Deepwater Horizon deep-sea oil plume. *ISME J* **6:** 451-460.

Mann EL, Chisholm SW. (2000). Iron limits the cell division rate of *Prochlorococcus* in the eastern equatorial Pacific. *Limnol Oceanogr* **45:** 1067-1076.

Martiny AC, Kathuria S, Berube PM. (2009). Widespread metabolic potential for nitrite and nitrate assimilation among *Prochlorococcus* ecotypes. *Proc Natl Acad Sci USA* **106:** 10787-10792.

Mason OU, Nakagawa T, Rosner M, Van Nostrand JD, Zhou J. (2010). First investigation of the microbiology of the deepest layer of ocean crust. *PLoS ONE* **5**.

Mayer JA, Taylor FJR. (1979). Virus which lyses the marine nanoflagellate *Micromonas pusilla*. *Nature* **281:** 299-301.

Mella-Flores D, Mazard S, Humily F, Partensky F, Mahé F, Bariat L *et al.* (2011). Is the distribution of *Prochlorococcus* and *Synechococcus* ecotypes in the Mediterranean Sea affected by global warming. *Biogeosciences* **8:** 2785-2804.

Milligan KLD, Cosper EM. (1994). Isolation of virus capable of lysing the brown tide microalga, *Aureococcus anophagefferens*. *Science* **266:** 805-807.

Moisander PH, Beinart RA, Hewson I, White AE, Johnson KS, Carlson CA *et al.* (2010). Unicellular cyanobacterial distributions broaden the oceanic N_2 fixation domain. *Science* **327:** 1512-1514.

Moisander PH, Morrison AE, Ward BB, Jenkins BD, Zehr JP. (2007). Spatial-temporal variability in diazotroph assemblages in Chesapeake Bay using an oligonucleotide *nifH* microarray. *Environ Microbiol* **9:** 1823-1835.

Moisander PH, Shiue L, Steward GF, Jenkins BD, Bebout BM, Zehr JP. (2006). Application of a *nifH* oligonucleotide microarray for profiling diversity of N_2-fixing microorganisms in marine microbial mats. *Environ Microbiol* **8:** 1721-1735.

Moisander PH, Zhang R, Boyle EA, Hewson I, Montoya JP, Zehr JP. (2012). Analogous nutrient limitations in unicellular diazotrophs and *Prochlorococcus* in the South Pacific Ocean. *ISME J* **6:** 733-744.

Moore JK, Donney SC, Lindsay K, Mahowald N, Michaels AF. (2006). Nitrogen fixation amplifies the ocean biogeochemical response to decadal timescale variations in mineral dust deposition. *Tellus: Series B* **58**: 560-572.

Moore LR, Post AF, Rocap G, Chisholm SW. (2002). Utilization of different nitrogen sources by the marine cyanobacteria, *Prochlorococcus* and *Synechococcus*. *Limnol Oceanogr* **47**: 989-996.

Morris RM, Longnecker K, Giovannoni SJ. (2006). *Pirellula* and OM43 are among the dominant lineages identified in an Oregon coast diatom bloom. *Environ Microbiol* **8:** 1361-1370.

Mosier AC, Francis CA. (2011). Determining the distribution of marine and coastal ammonia-oxidizing archaea and bacteria using a quantitative approach In: Klotz MG (ed) *Methods in Enzymology: Research on Nitrification and Related Processes,* Part A, vol. 486. Elsevier Academic Press: San Diego, CA, USA, pp 205-221.

Mulholland MR, Bronk D, Capone DG. (2004). Dinitrogen fixation and release of ammonium and dissolved organic nitrogen by *Trichodesmium* IMS101. *Aquat Microb Ecol* **37:** 85-94.

Mulholland MR, Capone DG. (2001). Stoichiometry of nitrogen and carbon utilization in cultured populations of *Trichodesmium* IMS101: Implications for growth. *Limnol Oceanogr* **46:** 436-43.

Neidhardt FC, Umbarger HE. (1996). Chemical composition of *Escherichia coli*. In: Neidhardt FC, Curtiss III R, Ingraham JL, Lin ECC, Low KB, Magasanik B, Reznikoff WS *et al*. (eds). *Escherichia coli* and *Salmonella*: *Cellular and Molecular Biology*. 2nd edn ASP Press: Washington, DC, USA, pp 13–16.

Orchard ED, Webb EA, Dyhrman S. (2009). Molecular analysis of the phosphorus starvation response in *Trichodesmium* spp. *Environ Microbiol* **11:** 2400-2411.

Paerl RW, Johnson KS, Welsh RM, Worden AZ, Chavez FP, Zehr JP. (2011). Differential distributions of *Synechococcus* subgroups across the California Current System. *Front Microbiol* **2**: 59.

Paerl RW, Turk KA, Beinart RA, Chavez FP, Zehr JP. (2012). Seasonal change in the abundance of *Synechococcus* and multiple distinct phylotypes in Monterey Bay determined by *rbcL* and *narB* quantitative PCR. *Environ Microbiol* **14:** 580-593.

Palenik B, Ren QH, Dupont CL, Myers GS, Heidelberg JF, Badger JH *et al.* (2006). Genome sequence of *Synechococcus* CC9311: Insights into adaptation to a coastal environment. *Proc Natl Scad Sci USA* **103:** 13555-13559.

Post AF, Penno S, Zandbank K, Paytan A, Huse S, Mark Welch D. (2011). Long term seasonal dynamics of *Synechococcus* population structure in the Gulf of Aqaba, Northern Red Sea. *Front Microbiol* **2**: 131.

Poulsen N, Kröger N. (2005). A new molecular tool for transgenic diatoms. *FEBS J* **272:** 3413-3423.

Rhee SK, Liu XD, Wu LY, Chong SC, Wan XF, Zhou JZ. (2004). Detection of genes involved in biodegradation and biotransformation in microbial communities by using 50-mer oligonucleotide microarrays. *Appl Environ Microbiol* **70:** 4303-4317.

Rich VI, Konstantinidis K, DeLong EF. (2008). Design and testing of 'genome-proxy' microarrays to profile marine microbial communities. *Environ Microbiol* **10:** 506-521.

Rich VI, Pham VD, Eppley J, Shi Y, DeLong EF. (2011). Time-series analyses of Monterey Bay coastal microbial picoplankton using a ‘genome proxy’ microarray. *Environ Microbiol* **13:** 116-134.

Rinta-Kanto JM, Buergmann H, Gifford SM, Sun S, Sharma S, del Valle DA *et al.* (2011). Analysis of sulfur-related transcription by *Roseobacter* communities using a taxon-specific functional gene microarray. *Environ Microbiol* **13:** 453-467.

Scanlan DJ, Bourne JA, Mann NH. (1996). A putative transcriptional activator of the Crp/Fnr family from the marine cyanobacterium *Synechococcus* sp. WH7803. *J Appl Phycol* **8:** 565-567.

Scanlan DJ, Ostrowski M, Mazard S, Dufresne A, Garczarek L, Hess WR *et al.* (2009). Ecological genomics of marine picocyanobacteria. *Microbiol Mol Bio Rev* **73**: 249-299.

Scanlan DJ, West NJ. (2002). Molecular ecology of the marine cyanobacterial genera. *FEMS Microbiol Ecol* **40**: 1-12.

Sebastian M, Ammerman JW. (2009). The alkaline phosphatase PhoX is more widely distributed in marine bacteria than the classical PhoA. *ISME J* **3:** 563-572.

Sebastian M, Gasol JM. (2013). Heterogeneity in the nutrient limitation of different bacterioplankton groups in the Eastern Mediterranean Sea. *ISME J 7*: 1665-1668.

Shi YM, Tyson GW, DeLong EF. (2009). Metatranscriptomics reveals unique microbial small RNAs in the ocean's water column. *Nature* **459:** 266-269.

Singh AK, McIntyre LM, Sherman LA. (2003). Microarray analysis of the genome-wide response to iron deficiency and iron reconstitution in the cyanobacterium Synechocystis sp PCC 6803. *Plant Physiology* **132:** 1825-1839.

Smith MW, Herfort L, Tyrol K, Suciu D, Campbell V, Crump BC *et al.* (2010). Seasonal changes in bacterial and archaeal gene expression patterns across salinity gradients in the Columbia River coastal margin. *PLoS One* **5**: e13312.

Smyth GK. (2005). Limma: Linear models for microarray data. In: Gentleman R, Carey V, Dudoit S, Irizarry R and Huber W (eds). *Bioinformatics and Computational Biology Solutions using R and Bioconductor*. Springer: New York, NY, USA, pp 397-420.

Song BK, Ward BB. (2004). Molecular characterization of the assimilatory nitrate reductase gene and its expression in the marine green alga *Dunaliella tertiolecta* (Chlorophyceae). *J Phycol* **40:** 721-731.

Stuart RK, Dupont CL, Johnson DA, Paulsen IT, Palenik B. (2009) Coastal strains of marine *Synechococcus* exhibit increased tolerance to copper shock and a distinctive transcriptional response relative to open ocean strains. *Appl Environ Microbiol* **75**: 5047-5057.

Sun S, Chen J, Li W, Altinatas I, Lin A, Peltier S *et al.* (2011). Community cyberinfrastructure for Advanced Microbial Ecology Research and Analysis: the CAMERA resource *Nucl Acids Res* **39:** D546-D551.

Sun J, Steindler L, Thrash JC, Halsey KH, Smith DP, Carter AE *et al.* (2011). One Carbon Metabolism in SAR11 Pelagic Marine Bacteria. *PLoS One* **6**: e23973. doi:10.1371/journal.pone.0023973.

Sunda WG, Swift DG, Huntsman SA. (1991). Low iron requirement for growth in oceanic phytoplankton. *Nature* **351:** 55-57.

Sylvan J, Quigg A, Tozzi S, Ammerman JW. (2007). Eutrophication-induced phosphorus limitation in the Mississippi river plume: evidence from fast repetition rate fluorometry. *Limnol Oceanogr* **34:** 2679-2685.

Sylvan JB, Quigg A, Tozzi S, Ammerman JW. (2011). Mapping phytoplankton community physiology on a river impacted continental shelf: testing a multifaceted approach. *Estuaries and Coasts* **34:** 1220-1233.

Taroncher-Oldenburg G, Griner EM, Francis CA, Ward BB. (2003). Oligonucleotide microarray for the study of functional gene diversity in the nitrogen cycle in the environment. *Appl Environ Microbiol* **69:** 1159-1171.

Tetu SG, Brahamsha B, Johnson DA, Tai V, Phillippy K, Palenik B *et al.* (2009). Microarray analysis of phosphate regulation in the marine cyanobacterium *Synechococcus* sp WH8102. *ISME J* **3:** 835-849.

Thompson AW, Huang K, Saito MA, Chisholm SW. (2011). Transcriptome response of high- and low-light-adapted *Prochlorococcus* strains to changing iron availability. *ISME J* **5:** 1580-1594.

Tiquia SM, Wu L, Chong SC, Passovets S, Xu D, Xu Y *et al*. (2004). Evaluation of 50-mer oligonucleotide arrays for detecting microbial populations in environmental samples. *BioTechniques* **36:** 664-675.

Tolonen AC, Aach J, Lindell D, Johnson ZI, Rector T, Steen R *et al.* (2006). Global gene expression of *Prochlorococcus* ecotypes in response to changes in nitrogen availability. *Mol Syst Biol* **2**: 53.

Tusher VG, Tibshirani R, Chu G. (2001). Significance analysis of microarrays applied to the ionizing radiation response. *Proc Natl Acad Sci USA* **98:** 5116-5121.

Vogel J, Axmann IM, Herzel H, Hess WR. (2003). Experimental and computational analysis of transcriptional start sites in the cyanobacterium *Prochlorococcus* MED4. *Nucl Acids Res* **31:** 2890-2899.

Ward BB, Eveillard D, Kirshtein JD, Nelson JD, Voytek MA, Jackson GA. (2007). Ammonia-oxidizing bacterial community composition in estuarine and oceanic environments assessed using a functional gene microarray. *Envirol Microbiol* **9:** 2522-2538.

Watkins-Brandt KS, Letelier RM, Spitz YH, Church MJ, Böttjer, D, White AE. (2011). Addition of inorganic or organic phosphorus enhances nitrogen and carbon fixation in the oligotrophic North Pacific. *Mar Ecol Prog Ser* **432**: 17-29.

Wawrik B, Boling WB, Van Nostrand JD, Xie J, Zhou J, Bronk DA. (2012). Assimilatory nitrate utilization by bacteria on the West Florida Shelf as determined by stable isotope probing and functional microarray analysis. *FEMS Microbiol Ecol* **79:** 400-11.

Webb EA, Moffett JW, Waterbury JB. (2001). Iron stress in open-ocean cyanobacteria (*Synechococcus*, *Trichodesmium*, and *Crocosphaera* spp.): Identification of the IdiA protein. *Appl Environ Microbiol* **67:** 5444-5452.

White AE, Spitz YH, Letelier RM. (2007). What factors are driving summer phyto-plankton blooms in the North Pacific Subtropical Gyre? *J Geophys Res* **112**:C12006.

Wu L, Kellogg L, Devol AH, Tiedje JM, Zhou J. (2008). Microarray-based characterization of microbial community functional structure and heterogeneity in marine sediments from the Gulf of Mexico. *Appl Environ Microbiol* **74:** 4516-4529.

Wu LY, Thompson DK, Li GS, Hurt RA, Tiedje JM, Zhou JZ. (2001). Development and evaluation of functional gene arrays for detection of selected genes in the environment. *Appl Environ Microbiol* **67:** 5780-5790.

Xie J, He Z, Liu X, Liu X, Van Nostrand JD, Deng Y *et al.* (2011). GeoChip-based analysis of the functional gene diversity and metabolic potential of microbial communities in acid mine drainage. *Appl Environ Microbiol* **2011:** 3.

Yergeau E, Kang S, He Z, Zhou J, Kowalchuk GA. (2007). Functional microarray analysis of nitrogen and carbon cycling genes across an Antarctic latitudinal transect. *ISMe J* **1:** 163-179.

Zehr JP, Montoya JP, Jenkins BD, Hewson I, Mondragon E, Short CM *et al.* (2007). Experiments linking nitrogenase gene expression to nitrogen fixation in the North Pacific subtropical gyre. *Limnol Oceanogr* **52:** 169-183.

Zhou A,He Z,Qin Y, Lu Z, Deng Y,Tu Q *et al.* (2013). StressChip as a high-throughput tool for assessing microbial community responses to environmental stresses. *Environ Science and Technol* **47**: 9841−9849.

Figures

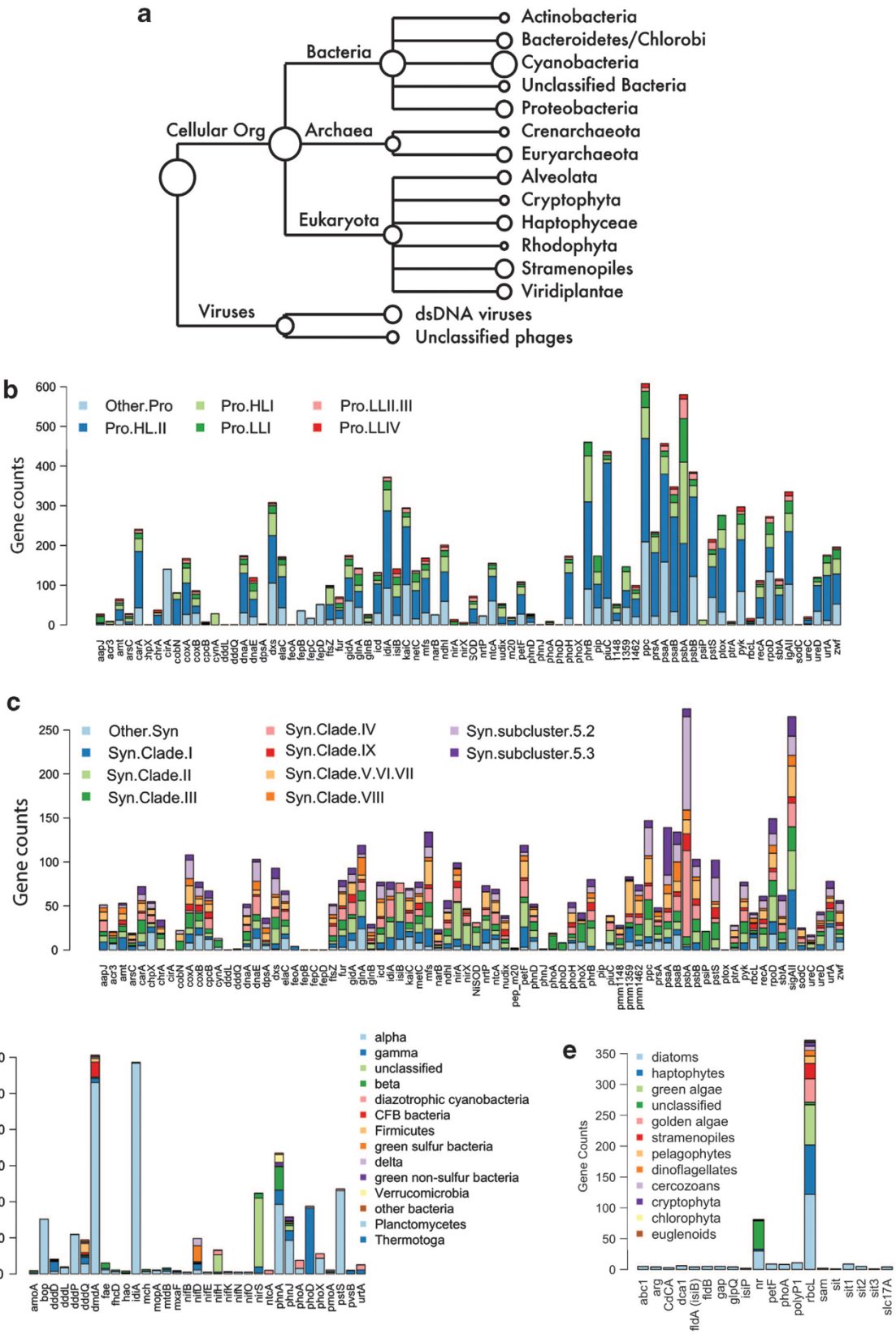
a
Actinobacteria
Bacteroidetes/Chlorobi
Cyanobacteria
Unclassified Bacteria
Proteobacteria
Crenarchaeota
Euryarchaeota
Alveolata
Cryptophyta
Haptophyceae
Rhodophyta
Stramenopiles
Viridiplantae
dsDNA viruses
Unclassified phages
Bacteria
Archaea
Eukaryota
Viruses
Cellular Org
b
Other.Pro
Pro.HL.II
Pro.HLI
Pro.LLI
Pro.LLII.III
Pro.LLIV
Gene counts
c
Other.Syn
Syn.Clade.I
Syn.Clade.II
Syn.Clade.III
Syn.Clade.IV
Syn.Clade.IX
Syn.Clade.V.VI.VII
Syn.Clade.VIII
Syn.subcluster.5.2
Syn.subcluster.5.3
Gene counts
d
alpha
gamma
unclassified
beta
diazotrophic cyanobacteria
CFB bacteria
Firmicutes
green sulfur bacteria
delta
green non-sulfur bacteria
Verrucomicrobia
other bacteria
Planctomycetes
Thermotoga
Gene counts
e
diatoms
haptophytes
green algae
unclassified
golden algae
stramenopiles
pelagophytes
dinoflagellates
cercozoans
cryptophyta
chlorophyta
euglenoids
Gene Counts

Figure 1. Microorganisms and viruses with genes targeted in the MicroTOOLs microarray. (a) Phylogenetic tree of all targeted genes with the number of genes reflected in the size of the circles. (b-e) Distribution of target genes in (b) *Prochlorococcus* spp. by clade, (c) *Synechococcus* spp by clade, d) bacteria other than *Prochlorococcus* and *Synechococcus* by phylogroups, e) Eukaryota by phylogroup. The details for gene distribution are in Table S1.

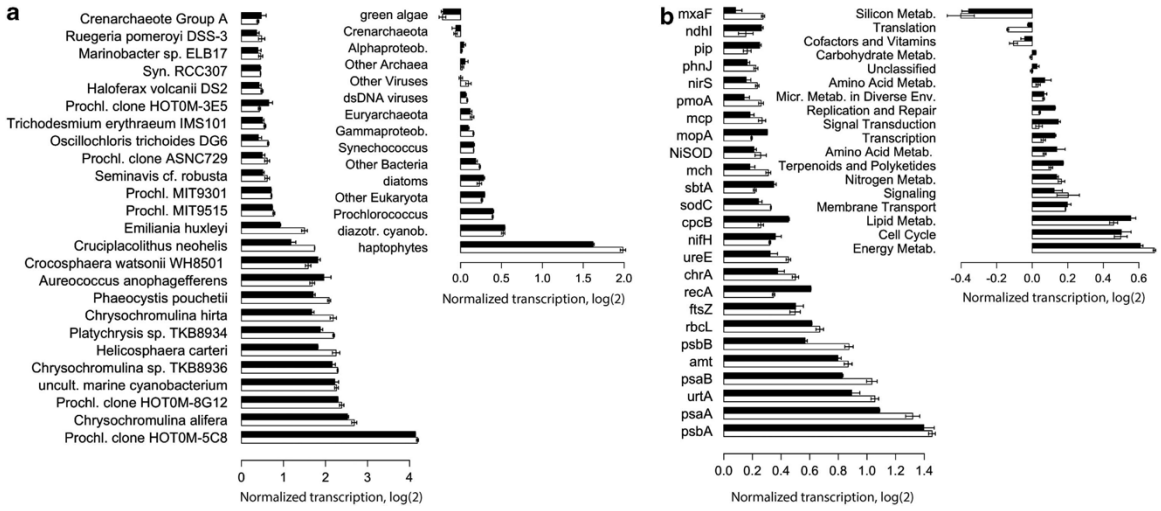

Figure 2. Average detected transcription by organisms (a) and by genes (b) normalized to the median transcription in each phylogroup in each sample. Only top 25 entries for each category are shown. Average transcription by phylogroups (a) and by gene pathways (b) are included as inserts. White and black represent P and Fe-amended treatments, correspondingly.

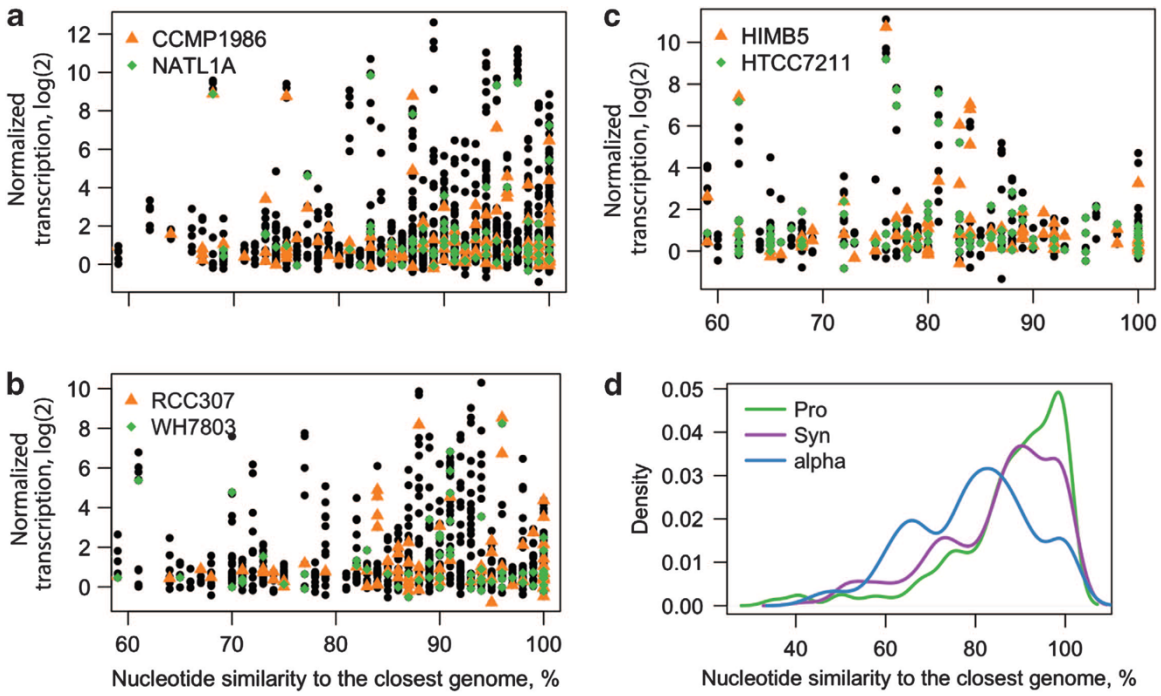

Figure 3. Transcription versus nucleotide similarity to the closest genome for *Prochlorococcus* spp (a) with CCMP1986- and NATL1A-like genotypes highlighted; *Synechococcus* spp (b), with RCC307- and WH7803-like genotypes highlighted; and Alphaproteobacteria (c) sequences, with HIMB5-and HTCC7211-like genotypes highlighted. Transcription was normalized to the median of each group in each sample, and only differentially transcribed genes are shown. d) Distribution of differentially transcribed genes by nucleotide similarity to the closest genome for *Prochlorococcus*, *Synechococcus*, and Alphaproteobacteria.

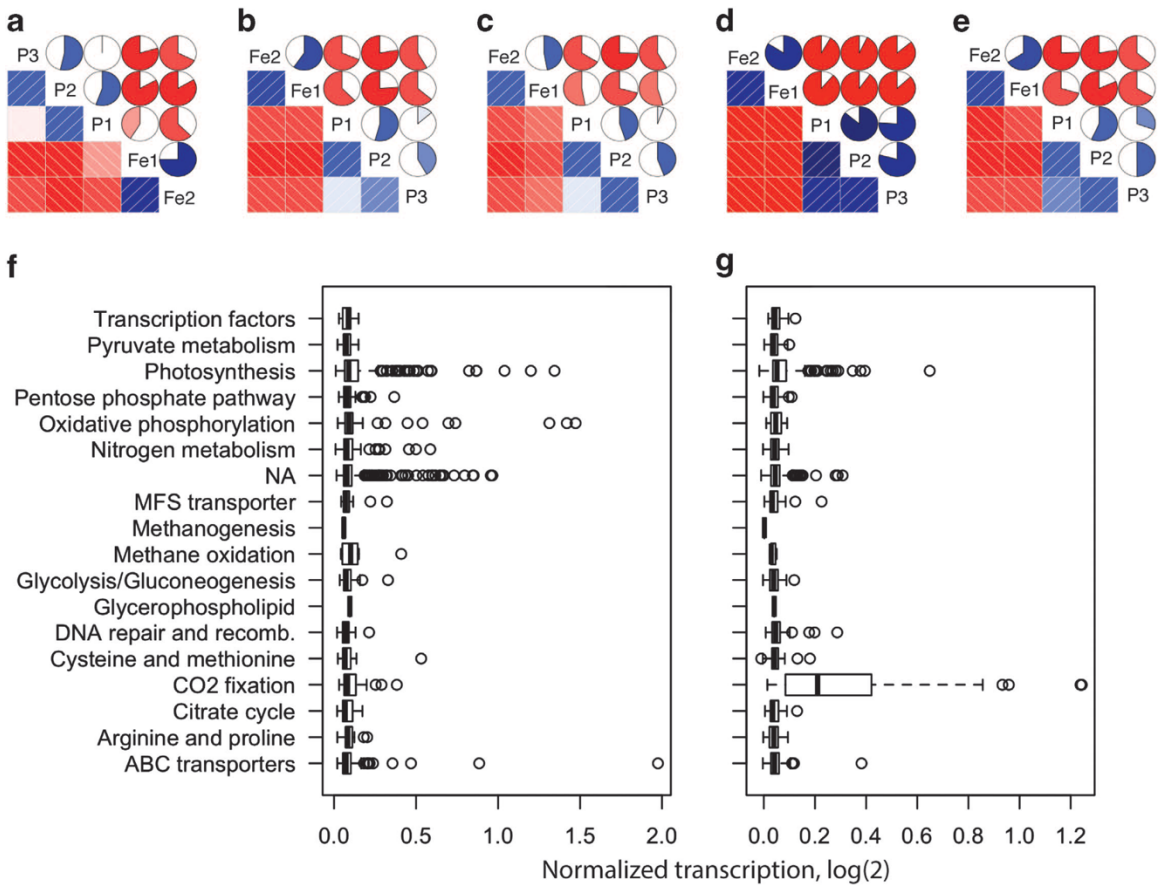

Figure 4. Correlograms for differentially transcribed genes from (a) Eukaryota (172 genes), (b) *Prochlorococcus* (2022 genes), (c) *Synechococcus* (1030 genes), (d) Energy metabolism (699 genes), (e) Nitrogen metabolism (472 genes). The low triangle from the principal diagonal line in each panel contains correlation cells for each pair of samples, and the upper triangle contains pie charts showing the strength of correlation. Blue and red represent positive and negative correlations, respectively, and color intensity reflects magnitude of the correlation. Columns and rows were reordered based on principal component to reflect samples that group together. Correlation coefficients for all correlograms are in Supplementary Material Table S4. Transcription of genes up-regulated in the Fe amendment (f) and in the P-amendment (g) by KEGG pathway.

Transcription was normalized to the mean across samples. 'NA' category comprises of genes not assigned to a KEGG pathway.

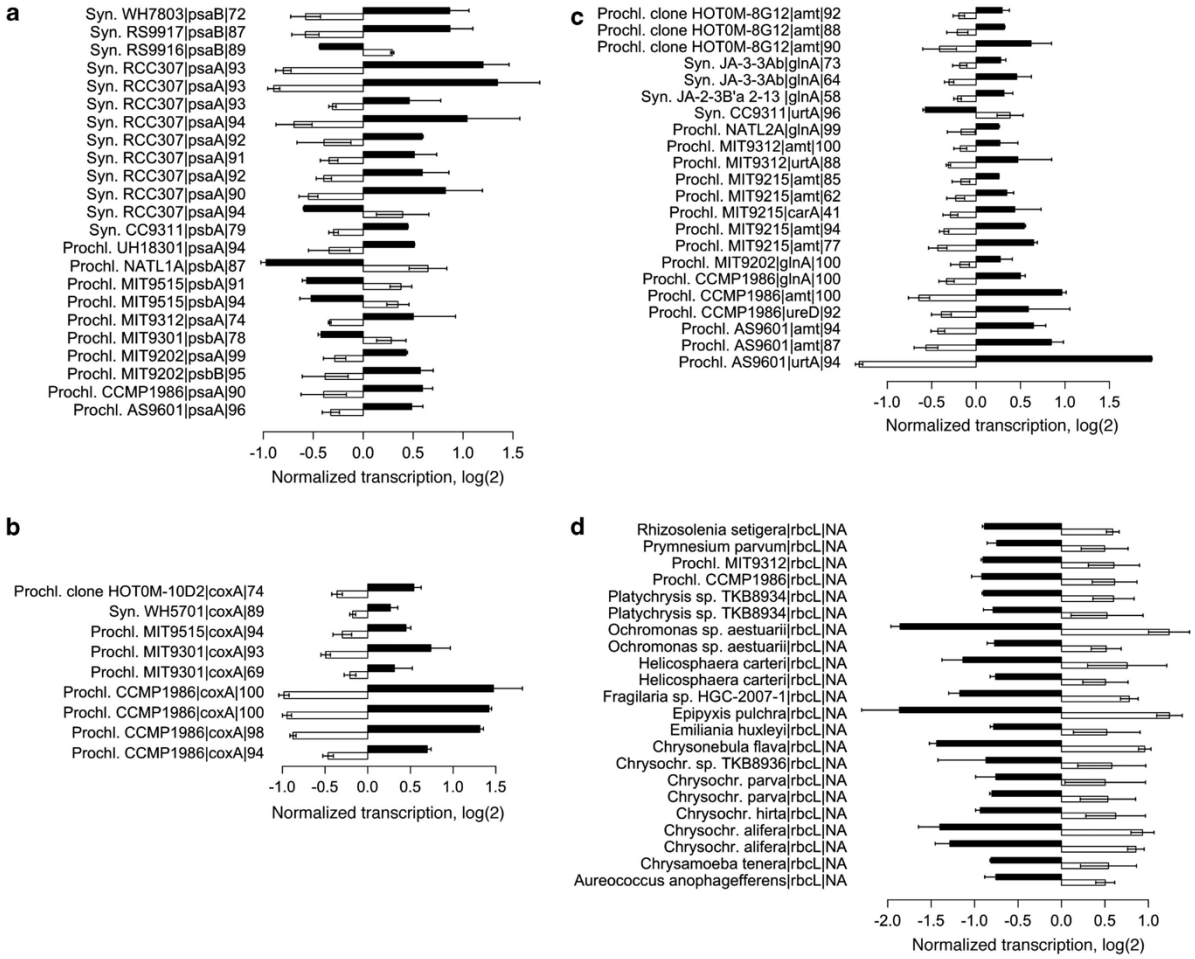

Figure 5. Normalized transcription for top differentially transcribed genes for (a) photosynthesis, (b) oxidative phosphorylation, (c) nitrogen metabolism and stress, and (d) carbon fixation. Transcription was normalized to the mean transcription across samples.

Tables

Table 2. Cyanobacterial abundances measured as gene copy numbers with qPCR and chlorophyll *a* and Fv/Fm measured with FRRF. The Tukey's significance test: ** 0.01, * 0.05. Cnt stands for Control.

	Treatments			t value		
	Cnt	**P**	**Fe**	**P-Cnt**	**Fe-Cnt**	**Fe-P**
Synechococcus, narB L-1	3.2±0.9E+03	6.7±1.3E+03	7.4±2.0E+04	0.22(n=6)	4.53**(n=6)	4.31**(n=6)
Ca. A. thalassa, *nifH* L-1	1.3±0.2E+03	3.2±0.6E+03	2.4±1.0E+04	0.26(n=6)	3.10*(n=6)	2.84*(n=6)
Crocosphaera, nifH L-1	4.3±0.6E+04	1.4±0.4E+05	3.2±1.1E+05	1.11(n=6)	3.19*(n=6)	2.09(n=6)
Fm	3.71±0.28	4.39±0.16	4.18±0.44	1.85(n=9)	1.13(n=9)	-0.61(n=9)
Fv/Fm	0.50±0.04	0.48±0.01	0.43±0.01	-1.07(n=9)	-2.77*(n=9)	-1.96(n=9)

Table 1. Marker genes targeted in the MicroTOOLs microarray.

Process	Gene	Annotation	Counts
Carbon metabolism	CdCA	cadmium containing Carbonic Anhydrase	3
	chpX	CO2 hydration protein ChpX	55
	dca1	delta carbonic anhydrase	6
	dxs	1-deoxy-D-xylulose-5-phosphate synthase	401
	fae	formaldehyde activating enzyme	30
	fhcD	formylmethanofuran-tetrahydromethanopterin formyltransferase	11
	gap	glyceraldehyde-3-phosphate dehydrogenase	5
	gidA	glucose inhibited division protein A	269
	icd	isocitrate dehydrogenase	209
	mch	methenyltetrahydromethanopterin cyclohydrolase	12
	mtdB	methylenetetrahydromethanopterin dehydrogenase	16
	mxaF	methanol dehydrogenase	9
	pmoA	methane monooxygenase	11
	ppc	phosphoenolpyruvate carboxylase	755
	prsA	ribose-phosphate pyrophosphokinase	282
	pyk	pyruvate kinase	374
	rbcL	RuBisCO	468
	sbtA	sodium-dependent bicarbonate transporter	177
	zwf	glucose-6-phosphate dehydrogenase	252
Cell cycle and Replication	CwatDRAFT_4045	Transposase CwatDRAFT_4045	1
	dnaA	replication initiation protein	226
	dnaE	DNA Polymerase III, alpha subunit	222
	elaC	ribonuclease Z	238
	ftsZ	cell division protein FtsZ	154
	kaiC	circadian clock protein KaiC	365
	pol	DNA polymerase	53
	recA	recombinase A	172
	rpoD	RNA polymerase sigma factor	143
	sigA	RNA polymerase sigma factor	279
DMSP metabolism	dddD	DMSP CoA transferase	41
	dddL	DMSP lyase	20
	dddP	DMSP lyase	110
	dddQ	DMSP lyase	97
	dmdA	dimethyl sulfoniopropionate demethylase	609
Energy metabolism	coxA	cytochrome c oxidase subunit I	275
	coxB	cytochrome c oxidase subunit II	163
	cpcB	phycocyanin, beta subunit	75
	hupS	Ni-Fe hydrogenase, small subunit HupS	3
	ndhI	NADH dehydrogenase subunit I	257
	petB	cytochrome b6f	265
	psaA	photosystem I P700 chlorophyll a apoprotein A1	599
	psaB	photosystem I P700 chlorophyll a apoprotein A10	482
	psbA	photosystem II PsbA protein (D1)	854
	psbA1	photosystem II PsbA protein (D1)	2
	psbA2	photosystem II PsbA protein (D1)	2
	psbB	photosystem II PsbB protein (CP47)	491
Iron metabolism	abc1	ABC1 Superfamily Protein	5
	cirA	ferric iron-catecholate outer membrane transporter	140
	dpsA	ferritin-like diiron-binding domain	42
	feoA	ferrous iron transport protein A	4
	fepB	ABC-type Fe3+-hydroxamate transport system	35
	fepC	ABC-type cobalamin/Fe3+-siderophores transport	16
	fepD	Fe3+ siderophore transport system	51
	fldA (isiB)	Flavodoxin eukaryotic	4
	fldB	Flavodoxin	5
	fur	ferric transcriptional regulator	149
	idiA	iron (III) transporter	1035
	isiA	iron stress induced chlorophyll binding protein	3
	isiB	flavodoxin	217
	isiP	Iron stress induced protein	2
	pep_m20	possible Peptidase family M20/M25/M56	21
	petF	ferredoxin	236
	piuC	uncharacterized iron-regulated protein	476
	pmm1359	Predicted membrane protein, iron-stress responsive	229
	pvsB	vibrioferrin biosynthesis protein PvsB	10
	sam	SAM-methyltransferase	2
Nitrogen metabolism	aapJ	polar amino acid ABC transporter	79
	amoA	ammonia oxidation	29
	amt	ammonium transporter	118
	arg	N-Acetyl Transferase	4
	carA	carbamoyl-phosphate synthase	313
	cynA	cyanate transporter	41
	glnA	glutamine synthetase	262
Nitrogen metabolism	glnB	N regulatory protein P-II	55
	hao	hydroxylamine oxidoreductase	9
	metC	Cystathionine beta-lyase family protein involved in Al resistance	213
	narB	assimilatory nitrate reductase in bacteria	48
	nifB	nitrogenase cofactor biosynthesis protein NifB	5
	nifD	nitrogenase reductase	98
	nifE	nitrogenase MoFe cofactor biosynthesis protein NifE	5
	nifH	nitrogenase iron protein NifH	66
	nifK	nitrogenase molybdenum-iron protein beta chain	6
	nifN	nitrogenase molybdenum-iron cofactor biosynthesis protein NifN	5
	nifO	nitrogenase-associated protein NifO	5
	nifX	nitrogenase molybdenum-iron protein NifX	1
	nirA	ferredoxin-nitrite reductase	113
	nirS	dissimilatory nitrite reductase	224
	nirX	homeobox domain, in the nirA operon	53
	NR	assimilatory nitrate reductase	81
	nrtP	nitrate transporter	95
	ntcA	N limitation transcriptional regulator	241
	slc17A	amino acid transporter	4
	Tery_2117	hypothetical protein, expressed as nif	1
	Tery_2900	hypothetical protein, expressed as nif	3
	Tery_4333	LysR family transcriptional regulator, expressed as nif	3
	ure	urea transporter	4
	ureA	urease alpha subunit	5
	ureB	urease beta subunit	4
	ureC	urease	40
	ureD	urease accessory protein UreD	167
	ureE	urease accessory protein UreE	4
	ureF	urease accessory protein UreF	6
	ureG	urease accessory protein UreG	7
	ureH	urease accessory protein UreH-like protein	1
	ureX	urease subunit	1
	urtA	urea ABC transporter, substrate binding protein	292
Other metabolisms	bop	proteorhodopsin	152
	chrA	chromate transporter	71
	cobN	cobaltochelatase CobN	104
	mopA	Heme binding region from putitive Mn-oxidase	11
Other stresses	mfs	multidrug efflux transporter, proline/betaine transporter	302
	NiSOD	putative nickel-containing superoxide dismutase precursor	103
	NUDIX	nudix hydrolase	92
	phrB	DNA photolyase	540
	pip	proline iminopeptidase	173
	pmm1148	EF-1 guanine nucleotide exchange	80
	pmm1462	conserved hypothetical protein PMM1462	173
	ptox	plastoquinol terminal oxidase	276
	sigAII	type II alternative sigma-70 family RNA polymerase sigma factor	600
	sodC	Cu-Zn superoxide dismutase	25
Phosphorus metabolism	acr3	arsenite transport (efflux)	31
	arsC	arsenate reductase	47
	glpQ	glycerophosphoryl diester phosphodiesterase	4
	phnA	phosphonoacetate hydrolase	335
	phnD	phosphonate transporter	81
	phnJ	phosphonate lyase	159
	phoA	alkaline phosphatase, Zn2+ binding	73
	phoD	alkaline phosphatase	195
	phoH	P stress inducible protein	227
	phoU	transcriptional regulator, phosphate transport system protein	1
	phoX	alkaline phosphatase, Ca2+ binding	98
	polyP1	poly-phosphate accumulation	11
	psiP	highly expressed under low P	33
	pstS	phosphate transporter	552
	ptrA	possible P transcriptional regulator	36
	sqdB	sulfolipid biosynthesis protein	1
Silicon transport	sit	silicon transporter	1
	sit1	silicon transporter	9
	sit2	silicon transporter	5
	sit3	silicon transporter	1
Viral genes	dnaPol	viral DNA polymerase	54
	g20	Viral capsid assembly protein g20	30
	gp23	viral major capsid proteins	6
	mcp	viral major capsid proteins	15
	RdRp	RNA-dependent RNA pol	22