The Molecular Biology of Biogeochemistry: Using Molecular Methods to Link Ocean Chemistry with Biological Activity
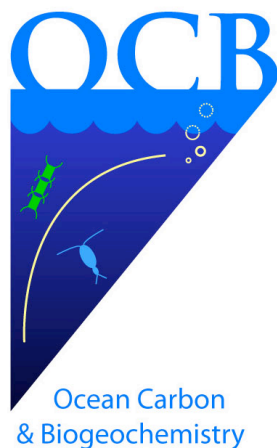
November 8-10, 2010 (Los Angeles, CA)

http://www.whoi.edu/sites/molbiogeochem

An Ocean Carbon and Biogeochemistry Scoping Workshop

Authors: James Moffett (USC), Eric Webb (USC), Bob Anderson (Columbia), Ginger Armbrust (UW), Kevin Arrigo (Stanford), Bess Ward (Princeton), Keith Moore (UCI), Benjamin Van Mooy (WHOI), Jed Fuhrman (USC)

Abstract

Currently, two large-scale, sectional survey programs in geochemistry are being conducted on a global scale: CLIVAR and GEOTRACES. CLIVAR is a US effort, while GEOTRACES is a collaborative effort involving many nations. Such survey programs have been conducted infrequently in the past. The goal of this workshop was to explore whether a sectional-survey approach would also be important for marine microbiology and if so, whether the activities of the existing programs, especially GEOTRACES, could complement a new program in some way. The over-arching theme that emerged from the conference is that our basic ignorance of marine microbial biogeography prevents us from answering many important questions in biogeochemistry, particularly involving carbon cycling and ecosystem response to climate change. Surface to seafloor, cross - basin, sectional surveys, coupled with physical and chemical data including the core parameters in GEOTRACES, would provide a means to identify microbial provinces or biomes and identify the physical and chemical parameters that define their boundaries and connections with each other. This information is necessary to understand how microbial biodiversity and biogeography influence marine biogeochemistry. Such knowledge would greatly enhance our understanding of carbon and energy flow within and between these biomes, and how their boundaries might change as a result of climate change.

Introduction

Biogeochemical cycles in the water column of the oceans are largely driven by unicellular organisms, which are controlled in part by their physical and chemical environment. This "bottom-up" perspective has been the foundation of a highly productive interaction for over three decades between marine microbiologists and chemists who study biogeochemical cycles. The adaptation of microbes to harsh chemical regimes associated with intense competition for scarce nutrients, micronutrients, and labile reservoirs of organic carbon has been studied in detail at the molecular level, providing new insights about global-scale processes, however these studies are often only focused on one to handful of microbial genera.

The past decade has seen the development of a number of global-scale, coupled ocean circulation-biogeochemical models for use in studies of the biogeochemical cycles and climate projection. These Dynamic Green Ocean Models (DGOMs) incorporate quantitative, mechanistic information about growth rates and micro- and macronutrient utilization for key phytoplankton taxa to study the spatial and temporal variability of primary production in the oceans and to predict how the ocean carbon cycle may change in the future (Le Quere et al., 2005; Moore et al., 2004). These models have revealed that the oceans can be divided into distinct biogeochemical provinces based on growth-limitation of key taxa by specific factors. The boundaries of these provinces are defined in the model by the simulated chemical and physical environment, and they are subject to significant temporal variability as these variables are influenced by climate change. Many DGOMs now include multiple phytoplankton functional groups, or taxa, explicitly, and there is great interest in incorporating additional biological complexity (additional phytoplankton, zooplankton groups, and heterotrophic bacteria including those associated with the oxygen minimum zones). However, progress is currently hampered by the lack of data constraining microbial biogeography at large spatial scales. Validation of these models, and the generation of new data to refine them for future forecasts, is an important objective of global survey programs like GEOTRACES. Indeed, the major thrust in oceanography in this century has been the establishment of new programs like

GEOTRACES, CLIVAR and the Ocean Observing Initiative (OOI) to enhance our observational capabilities. While global observational programs like GEOSECS (http://iridl.ldeo.columbia.edu/SOURCES/.GEOSECS/) and WOCE (http://woce.nodc.noaa.gov/wdiu/) have been carried out in the past, the scope of current programs represents a significant shift from the individual investigator-led programs, which are inherently limited by sampling capacity. These new programs provide more holistic sampling that promise improved models of biogeochemical interconnections in the ocean. The holistic sampling approach used in the JGOFS campaigns has driven much of the progress in marine ecosystem model development during the past two decades.

A central rationale of many global programs is to understand the carbon cycle more clearly, especially with regard to biologically driven processes. However, biological measurements in these programs have been limited because traditional methods were not appropriate for the sectional collection of many small, discrete samples. Fortunately, recent advances in molecular biology provide tools that can be extremely compatible with high throughput sampling, and provide an unprecedented degree of information about the linkage between microbial biodiversity and global biogeochemical cycles.

Present state of knowledge

In the last three decades, our understanding of microbial diversity in the oceans has changed dramatically, increasing from a handful of easily cultured representatives to hundreds of finished genomes, local and global metagenomic surveys, and scores of new phylotypes with undefined phenotypes and functions. These changes in our baseline knowledge have been driven largely by two factors: 1) the recognized importance of microbes in oceanic biogeochemistry, and 2) technological sampling/analysis advances both in the field and in the laboratory. From the initial datasets using flow cytometry and direct microscopy counts (Chisholm et al., 1988; Jannasch and Jones, 1959; Waterbury et al., 1979; Ferguson and Rublee, 1976), we knew the abundances of microbes in the oceans (~10^6 bacteria, 10^5 *Prochlorococcus*, and 10^3 *Synechococcus* per mL) and how these total values changed in different regimes. However, missing in these early studies

was information pertaining to changes in the diversity and function of microbes present and the environmental factors that control this variability. While initial phylogenetic studies were focused primarily on defining 'who' was there, based on the pioneering 16S cloning and sequencing work, later studies began and are currently asking questions of function, association, and biogeochemical impact (reviewed in Giovannoni and Stingl, 2005). For example, work built on these early findings has also uncovered many new first principles for marine microbes, including the realization that: Members of the genera *Prochlorococcus* and *Synechococcus* are the most abundant phototrophs on the planet (Johnson et al., 2006), many microbes (in addition to photoautotrophs) may use energy from the sun in unexpected ways, i.e. proteorhodopsin or bacteriochlorophyll (Beja et al., 2000; Gómez-Consarnau et al., 2007; Kolber et al., 2001; Moran and Miller, 2007; Venter et al., 2004), marine Archaea are abundant at depth (Fuhrman and Ouverney, 1998; Karner et al., 2001) and are likely to be autotrophic (Kuypers et al., 2001; Pearson et al., 2001; Schleper et al., 2005; Wuchter et al., 2006), nitrification occurs in both Eubacteria and Archaea (Könneke et al., 2005; Francis et al., 2005), and niche adaptation in the ocean can be biogeochemically controlled (Martiny et al., 2006; Rusch et al., 2010). Many other novel findings in previous decades have been reviewed (Giovannoni and Stingl, 2005; DeLong and Karl, 2005), thus we are truly in the midst of a 'microbial revolution' in biological oceanography. Unfortunately, in many cases, the linkages between old findings as well as these new ones and biogeochemistry are poorly constrained.

The combined application of modern techniques (e.g., quantitative PCR and/or metagenomics) with classical methods (e.g., flow cytometry and/or direct counts) to
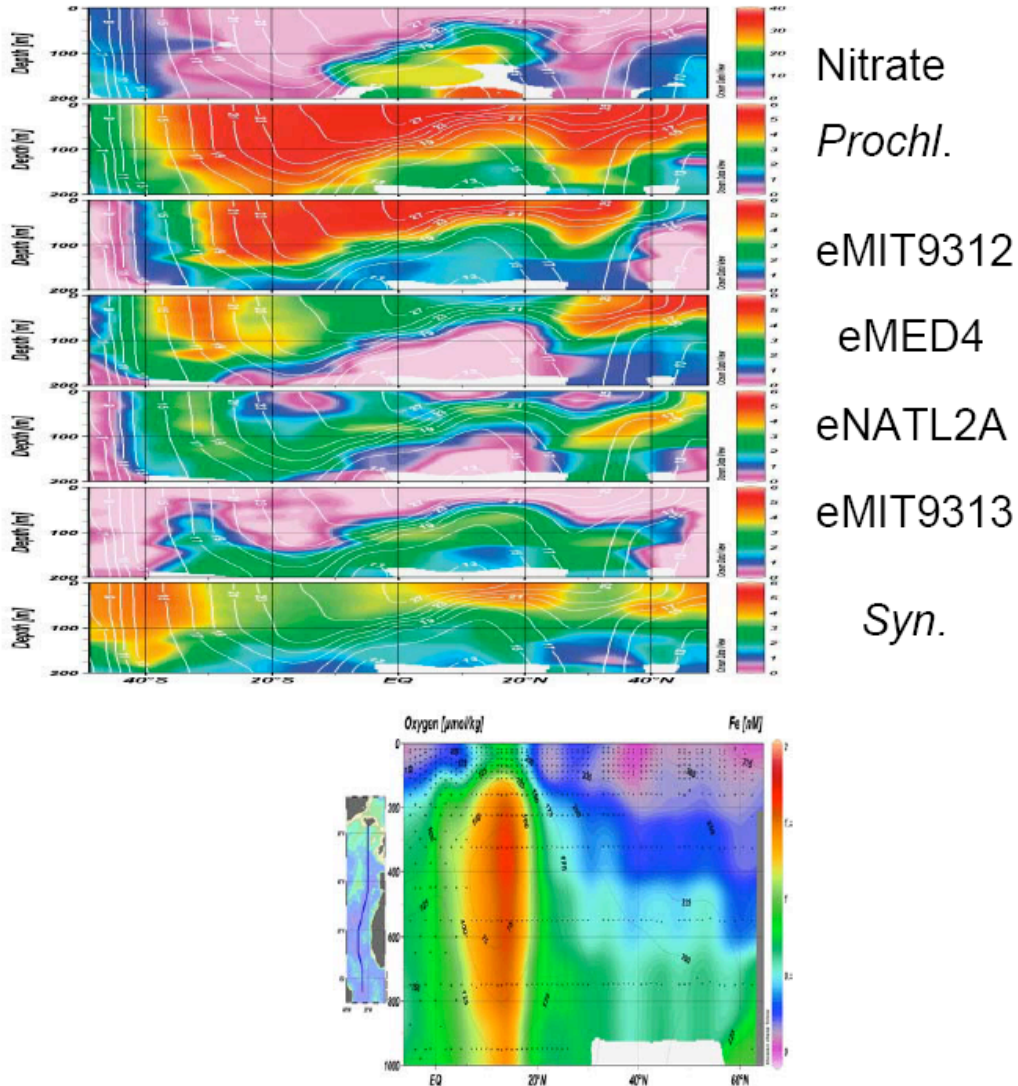

Figure 1. Top. Distribution of *Prochlorococcus* strains on a meridional section of the Atlantic Ocean (inset). From Johnson et al. (2006). Bottom: Dissolved Fe obtained on a back-to-back section. From Measures et al. (2008).

environmental samples has allowed more refined questions to be asked. For example, recent work coupling flow cytometry with RT-qPCR has permitted oceanic surveys of specific groups, including a meridional section in the North and South Atlantic Oceans, which showed that there is more *Prochlorococcus* diversity present than the existing

qPCR primers can account for and revealed unprecedented connections between *Prochlorococcus* and its physical and chemical environment (Figure 1). Indeed, the qPCR work explored by Chisholm and coworkers (e.g., Johnson et al., 2006) represents a useful model for how molecular biological, physiological, and geochemical data can be assimilated into the dataset as a whole when the target organisms are well defined. Figure 1 also shows dissolved Fe data obtained on a back-to back section on the same vessel. Interesting correlations are observed between *Synechococcus* and Fe, but the comparison would have been more valuable on a single cruise.

Alternatively, metagenomics and metatranscriptomics can be applied to ascribe characterized physiologies and phenotypes to uncultured phylotypes; concomitant physio-chemical measurements have the potential to provide broader biogeochemical context to these 'omics' datasets and define the connections between the biological and chemical species. *It was argued at the workshop, that in the absence of* **coordinated, high-resolution biological and chemical measurements**, *the interconnections and relationships that drive biogeochemistry in the ocean (i.e. microbial activity) could remain obscured in independent data sets.*

Novel charge of this workshop

Workshop participants were drawn from the biogeochemistry and marine microbiology communities, who shared a common interest in ecology and carbon cycle-related problems. There was considerable collective experience in interdisciplinary work; indeed many participants had collaborated with each other on projects of varying sizes. However, this was the first meeting to address collaboration on a large scale in a global survey and a number of scientific and logistical issues were discussed that were new to many participants. GEOTRACES was used as a representative program for several reasons. Many of its science objectives are biogeochemical and could benefit from biological context that molecular tools provide, and many of the personnel involved also collaborate with microbiologists on other projects. Furthermore, the process by which GEOTRACES was developed as a concept, funded by NSF, and implemented, could be a useful model for other programs. Perhaps most importantly, consideration of the

implementation of GEOTRACES sections helped participants focus on the science
questions and tools that were really appropriate for sectional surveys.

The workshop addressed the following major issues:

- What are the overarching questions and highest priorities for molecular
  measurements that could be incorporated into a sectional survey program like
  GEOTRACES or CLIVAR?
- Is GEOTRACES appropriate or should a stand-alone program with different
  sampling frequencies or cruise tracks be developed?
- What measurements could be carried out now? In 5-10 years? Is there a rationale
  for creating an archive of samples for analyses envisioned to be feasible in the
  next 10 years?
- What approaches can we use to synthesize molecular biological and geochemical
  data sets? How can the data be useful to modelers? What is an appropriate format
  for intercalibration studies? How will standard be developed?

During the workshop, participants generated a list of important questions in the global
ocean carbon cycle that were linked to microbial ecology. Most of these questions are
familiar; they have been examined in many previous forums. *This workshop was unique
because it specifically addressed how these questions could be addressed using
parameters derived from a large, basin-scale survey program.*

Important Questions:

1. What is the relationship between the photoautotrophic community and
   heterotrophic community in the euphotic zone? How is this relationship
   influenced by chemical and physical parameters? How does it ultimately
   influence community structure, nutrient utilization and the f-ratio?

2. What is the relationship between heterotrophy and heterotroph community
   structure in the deep ocean (including the "twilight zone") and the overlying
   biogeochemical regime? How does this affect vertical scale of organic carbon

regeneration (e.g., the Martin curve) and, ultimately, the oceanic control of atmospheric CO_2 (e.g., Kwon et al., 2009)? What is the relative importance of silicate versus PIC-dominated phytoplankton systems in contributing to carbon export in the deep ocean?

3. How is climate change likely to influence upper ocean ecosystems? Will it be primarily through changing the boundaries of existing biogeochemical provinces or by establishing new, unprecedented ones.   An example would be high-pCO_2 oligotrophic surface waters. In the contemporary surface ocean, pCO_2 is only high in nutrient-rich upwelling regions.

4. What really controls the growth rates of diatoms in different biogeochemical provinces of the oceans? What defines the boundaries? What species-specific differences exist within this diverse group of organisms that confounds quantification of bottom-up controls?

5. What controls the inventory of fixed N in the ocean? How important is denitrification vs. N_2 fixation and how will climate change affect the relative importance?

6. What is the relative importance of anammox versus denitrification, especially in oxygen minimum zones (OMZs)? How does their relative importance affect how the overall loss of fixed nitrogen is determined by physical and chemical parameters? What factors control the balance between these processes today, and how might this balance be altered by climate change, with concomitant feedbacks that may include N_2O?

Each of these questions (and other ones that emerged during the discussion) has not only been discussed at previous conferences, but is actively being worked on by individual PIs. The charge for this group was to determine how a sectional survey approach involving

many investigators could address these questions within a common program, and whether it would be the best use of resources.

The case for microbial biogeography

Discussion immediately centered on our relative ignorance of the distribution of microbes in the ocean, and their relationship to physical and chemical parameters. A key question was whether the concept of microbial biogeography, becoming a valuable tool in many habitats (Martiny et al., 2006), could be applied to open ocean marine microbial communities in a rigorous way. While useful in the upper ocean, characterizing distinct provinces in the deep ocean, dominated by heterotrophy and chemoautotrophy, would also be useful. The DGOMs (Moore et al., 2004) predict the existence of biogeochemical provinces in the upper ocean where the physiology and species composition of primary producers is controlled by bottom-up processes. These provinces are analogous to those identified by Alan Longhurst in his study of Oceanic Biogeography, especially the 2007 edition, which includes iron limitation (Longhurst, 2007). If we could define biogeographic regimes, then we could investigate the chemical and physical factors that control their boundaries. Additionally, if we could predict how climate change influences those factors, then we could predict how the boundaries of such regimes might change in the future. Characterization of microbial biogeography in this way could only be accomplished through sectional surveys involving the coordination of many groups measuring biological and chemical parameters, with molecular biological measurements at the core. Biogeochemists are excited about functional genomics as a diagnostic tool to probe how key processes are affected by their physical and chemical environment. But these diagnostics are hard to interpret in the absence of detailed knowledge of community composition, microbial physiology, and biogeochemical rates. It would of course be very exciting if the distributions of phylogenetically unrelated taxa exhibited common boundaries (as many microbial ecologists expect to be the case). This could indicate common regulatory controls (e.g., a high Fe requirement) or complex synergistic interactions that lead to distinctive and predictable microbial communities (Fuhrman, 2009). Characterization of community structure was widely seen as a key prerequisite for determining microbial function within the context of detailed geochemistry. Therefore,

10

based on multiple discussions both in breakout groups and in plenary, participants concluded that linkages between microbial diversity and biogeochemical processes could be defined using modern methods in genomics.

Relationship to GEOTRACES

Characterization of microbial biogeography is an ambitious problem. A good case was made that the existing GEOTRACES program would not be able to fully accommodate a comprehensive characterization, so a stand-alone program was proposed. However, GEOTRACES was used as a starting point for discussion, and strong links between the proposed program and GEOTRACES are envisaged. What follows is a brief description of GEOTRACES, which illustrates its relevance to the discussion.

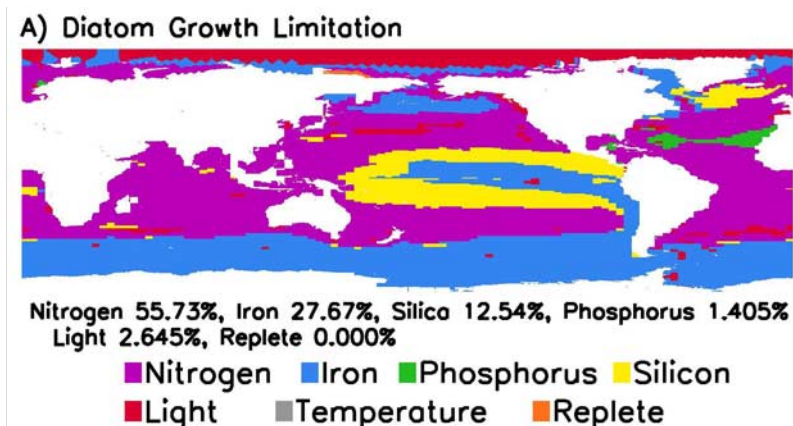GEOTRACES is an international survey program designed to measure a suite of trace elements and isotopes on a series of oceanographic sections. It was inspired by the GEOSECS program of the 1970s. Many of the core parameters – the key elements and isotopes measured on each cruise – are biologically essential and influence rates of primary production

Figure 2. Factor most limiting to growth rates is plotted for diatoms during summer months in each hemisphere (from Moore, 2004).

and the composition of phytoplankton assemblages. An important objective for GEOTRACES is to develop a database for biologically active metals that can be used to evaluate the importance of their variability to biogeochemical processes. Figure 2 (Moore, 2004) shows the distribution of factors limiting diatom growth, as predicted by this model. An important motivation to measure Fe in GEOTRACES (and CLIVAR) is to define the boundaries of these regions more accurately. With more accurate Fe data, it

should be possible to predict where Fe limitation is important. However, why not simply measure Fe limitation directly? Previously, such information could only be obtained through incubation experiments, which are not appropriate for a large survey program. Now, molecular tools give us the opportunity to examine Fe limitation and other important characteristics of phytoplankton regimes directly (Webb et al., 2007). This is a very good example of the value of a synthesis of geochemical and molecular data.

GEOTRACES sections are characterized by stations every 2.5 degrees, with samples collected from the surface to the seafloor. Core parameters include bioactive metals, other trace elements that are important paleotracers, and isotopes that are used to characterize the rates and mechanisms of important processes, like aerosol deposition and particle scavenging. The list of core parameters is shown in Table 1. Presently, the only core GEOTRACES "biological" parameters are chlorophyll and pigments. There is no time or space available on the vessel to support traditional process measurements. The massive data sets generated on these sections are archived through a GEOTRACES data management office in Britain, which works closely with the Biological and Chemical Oceanography Data Management Office (BCO-DMO) in Woods Hole, MA (USA). Before the first cruises were undertaken, a massive international intercomparison effort was undertaken for all of the core parameters over the course of two years.

Science objectives for the GEOTRACES program were formulated by a very inclusive process, with participation by a broad group representing multiple sub-disciplines of oceanography. Workshops were held for each ocean basin and sections were proposed, which were evaluated and prioritized by a scientific steering committee. For each basin, the workshop products were incorporated into a management proposal that was submitted to the Chemical Oceanography program. Individual proposals were then solicited for measurement of core parameters on the cruise, and evaluated by peer review in core program panels. While the overall GEOTRACES program is more exploratory than hypothesis-driven, each proposal defends specific hypotheses about the parameters proposed as well as the synthesis with the results from other parameters that is the foundation of the GEOTRACES mission.

OCB scoping workshop participants generally liked the GEOTRACES model, but there were two basic reasons why a stand-alone program was deemed necessary. First, it is quite clear that it is logistically impossible to accommodate the full sampling and personnel needs for extensive molecular biology on GEOTRACES cruises, which are already strained for space and ship time. A few samples are collected under the auspices of Biogeotraces (http://www.obs-vlfr.fr/GEOTRACES/index.php/science/biogeotraces), but on a small scale for independent PI projects. Secondly, the GEOTRACES sampling strategy, with a focus on high-resolution sampling throughout the water column, was seen by the workshop participants to have too much focus on the deep ocean and too little in the upper water column. Most importantly, some process work (e.g., rate measurement) was seen as desirable, and this required fewer stations to allow more time for these experiments. However, workshop participants recognized the value of the core GEOTRACES parameters, and a proposal to link a microbial cruise with a GEOTRACES cruise was formulated, and will be discussed below.

The many practical and logistical aspects of GEOTRACES, such as intercalibration, data synthesis and data management, were deemed highly relevant, although it was recognized that these topics are poorly developed in molecular biology. So they are developed further within this document.

Global surveys versus time-series stations and repeat surveys
Given finite resources for observational work, it is important to consider the relative merits of global surveys versus time-series stations. Time-series stations have proven to be exceedingly valuable to study temporal variability in biogeochemical and ecological processes on seasonal to decadal timescales. Sections only provide a snapshot in time for given locations. Global sections have been indispensable for our understanding of nutrients in the ocean – insight that would not be possible from the one-dimensional perspective of time-series station profiling. However, microbial community composition may be much more variable than nutrients alone. Should time-series stations receive a higher priority for support than sectional surveys? Are they more useful to characterize

ecosystem response to climate change, particularly if they could be located in areas where models suggest temperature changes will be most pronounced?

Much of the discussion at the workshop focused on characterizing microbial biogeography across a very broad range of marine biomes. These include areas like oxygen minimum zones, HNLC regions, highly productive marginal seas, and oligotrophic central gyres. Sampling such a wide range of regimes could only be accomplished with a dedicated survey program. But how useful would the information be without a good sense of temporal variability at each location within the survey? Participants made a strong case for the characterization of microbial biomes in surveys, accompanied by a vast array of physical and chemical data to establish causal relationships and provide information on spatial relationships between microbial, physical and chemical data that can be used to predict how the system might respond to climate change. Sarmiento et al. (2004) identified 6 major upper ocean biomes (Fig. 3), defined on the basis of physical parameters that control nutrient supply from underlying waters (vertical velocity, maximum winter mixed layer depth, and sea ice cover), and showed that these were very similar to the ecological provinces identified by Longhurst (2007). They showed that a principal effect of climate change on ecosystems will be a poleward shift in the boundaries of specific biomes. Such shifts can be predicted if our surveys provide information on the physical and chemical parameters that define biome boundaries.

A key difference between the biomes identified by Sarmiento et al (2004) and the regions defined in Moore et al. (2004) is that Fe was not considered in the former work. Climate change influences on dust will be significantly decoupled from effects on ocean circulation and mixing because of strong terrestrial influence. Neither study examined how the boundaries of these biomes might influence the characteristics of heterotrophic and chemoautotrophic communities in the deep ocean. With the exception of oxygen
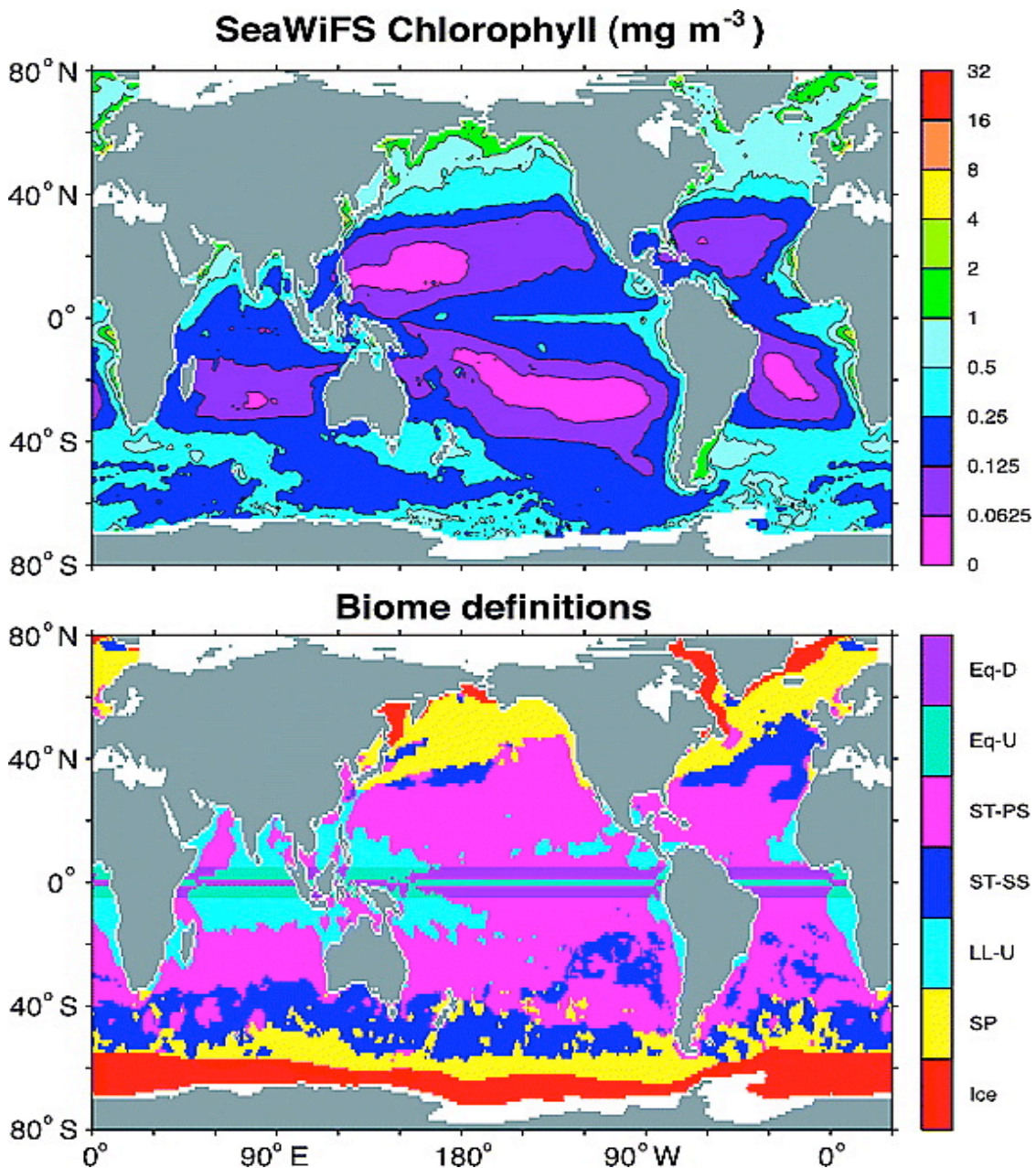
Figure 3. (a) Annual mean SeaWiFS chlorophyll from Yoder and Kennelly (2005) (online data set). (b) Biome classification scheme calculated using mixed layer depths obtained from observed density and from upwelling calculated from the wind stress divergence using observed winds. The equatorially influenced biome covers the area between 5°S and 5°N, and is colored a dirty light blue in areas where upwelling occurs (labeled "Eq-U" on the color bar) and dark pink in areas where downwelling occurs (labeled "Eq-D"). Outside of this band, the region labeled "Ice" (red) is the marginal sea ice biome, the region labeled "SP" (yellow) is the subpolar biome, the region labeled "LL-U" (light blue) is the low-latitude upwelling biome, the region labeled "ST-SS" (dark blue) is the seasonally mixed subtropical gyre biome, and the region labeled "ST-PS" (pink) is the permanently stratified subtropical gyre biome. From Sarmiento et al. (2004).

15

minimum zones, we have little basis for defining subsurface biomes on the basis of physics or chemistry. An important rationale for the whole-section approach is the ability to characterize sub-surface microbial biomes and connect them with conditions at the surface.

The principal measurements to be made on a section

Workshop participants preliminarily concluded that a holistic sampling scheme that married process and rate measurements with 'omic' work would be most compatible with the sectional approach. The consensus view was that filters for DNA and RNA would be collected for metagenomic and meta-transcriptomic analyses at each station, including multiple depths and size fractions. In addition to these sequencing-based approaches, microarrays already developed for the components of the nitrogen cycle could be a quick and efficient way to survey the 'known' diversity and activity (Moisander et al., 2007; Moisander et al., 2006; Taroncher-Oldenburg et al., 2003; Wu et al., 2001). Because of high volume requirements and concentration times, additional samples would be collected for proteomics and metabolomics only when possible (Table 1). However, even if limited compared to the DNA/RNA work, participants felt that these measurements would be an invaluable asset to the data analysis of the sectional cruise. Past work by Venter and colleagues on the Global Ocean Survey (GOS) (Rusch et al., 2010; Rusch et al., 2007; Yooseph et al., 2007) has shown the utility of metagenomic methodology, but many researchers have been disappointed by the absence of depth, chemical, and rate measurements from this ambitious study. The lack of this contextual information from the GOS cruise sampling has necessitated modeling the chemical components present during the cruise and thus *predicting* relationships between the biological community and bottom-up stressors (Rusch et al., 2010). In the proposed sectional cruise, the chemistry and biology would either be measured concomitantly or near-simultaneously from another ship. While the participants felt that the absolute list of measurements to be made on a section cruise should be the topic of another workshop, the major parameters likely to be measured are those listed in Table 1.

Sampling scheme	Parameter to be measured midday and midnight
Core stations	Metals, Nutrients (dissolved inorganic and dissolved organic), PAR, size fractionated Chla, POC, PON, Flow cytometry Bacterial/picoeuk counts, Microscopy counts of larger Euks, samples for microarrays, metagenomics and metatranscriptomics, Size spectrum from smallest viruses to largest protists.
'Rate' stations	Metabolomics, Proteomics, Nitrogen fixation, Denitrification, ^{15}N isotopic labeling for N cycling, Primary and Bacterial production, vitamin utilization, Stable isotope probing of nitrogen source utilization, Anammox, Manipulation experiments to test limitation

Table 1. Parameters to be measured on a sectional cruise (either by a second ship or on the GEOTRACES ship).

One major incentive for the 'omic' methodology was that the DNA/RNA for these analyses could easily be collected by a small, well-trained team, and then shared with the rest of the interested community, thus providing material for inter-comparisons and saving precious ship space for other groups to perform rate/process studies. Additionally, while seemingly expensive in the near term, the next-generation sequences approaches under development at this time were deemed the most cost effective in the long run, since these approaches offer both depth and breadth of sequence information, which will become an essential framework for parallel proteomic and metabolomic datasets. These methods are useful not only for linking known genes with known organisms, but also for mapping the occurrence of novel genotypes and uncharacterized biochemical pathways. Thus, it was recognized that the next generation techniques offer great potential for new discoveries with field samples (Beja et al., 2000; Rusch et al., 2010), and since the microbiology of the oceans is chronically under-sampled, they have the best 'novel finding':expense ratio. With this in mind, the workshop participants came up with the following priority list:

i) next generation metagenomics, transcriptomics (because these create a relatively unbiased digital archive that can be reinvestigated);

ii) tag rRNA sequencing (to survey the both the dominant and 'rare' biosphere);

iii) proteomics (leveraging the data obtained in the next generation sequencing efforts);

iv) tag sequencing to increase coverage of specific target genes linked to key biogeochemical pathways;

v) qPCR of these specific target genes;

vi) fosmids; and

vii) metabolomics, defined in the broadest and most ambitious sense to include all biochemistry left uncharacterized by the aforementioned RNA, DNA, and protein analyses.

Rationale for an Eastern Tropical South Pacific (ETSP) zonal section in conjunction with GEOTRACES

In 2011, a GEOTRACES Zonal Section across the North Atlantic will be concluded. A zonal section from Peru across the eastern tropical south Pacific has been proposed for Fall 2013. This cruise will include the Peru Upwelling region, the oxygen minimum zone, and the hydrothermal plume that extends westward from the East Pacific Rise. The workshop initially used this cruise as a "straw man" to encourage participants to examine in detail the organization and staging of a cruise in the near future with currently available technologies. However, considerable excitement was generated by the cruise track, since it spans a number of important microbial biomes, and the GEOTRACES Core parameters were seen as essential ancillary measurements. One proposal was to seek funding for a separate cruise, either simultaneously or back-to back, with the GEOTRACES function. While it is not feasible to establish a stand-alone program of the scope of GEOTRACES in that time frame, a pilot study that incorporates many of the elements discussed at the workshop might be possible.

Measurements for intercalibration

It is standard laboratory procedure for 'omic' work performed in the laboratory to include both technical and biological replicates. However, in mixed community metagenomic or transcriptomics analyses, while you can do replicate bottle incubations, true biological replicates are not possible, as the community can change in between sampling or be different due to micro-scale heterogeneity. Although measurement intercalibrations are not typically done in 'omic' field research, workshop participants thought there would be some value in limited intercalibrations to help with issues of blanks and contamination with foreign DNA, as well as extraction efficiencies and yields. If these sequencing efforts involved multiple PIs, then it was suggested that round robin calibration exercises be performed, similar to what was done in GEOTRACES and for Fe on the Sampling and Analysis of Fe (SAFe) cruise. However, in this case, a cruise would not be required; instead, these controls could be accomplished by all groups analyzing the 'omics' of some model marine microbes (both prokaryotic and eukaryotic) or some complex seawater sample as a reference standard. Important in this work would be the replication within as well as between laboratories. Alternatively, conducting all of the 'omics' work at a single analytical center (like Venter or JGI) would obviate the need for round robin intercalibration studies and minimize contamination concerns.

Defining appropriate normalization for transcriptomic and proteomic expression level analyses was also discussed. Ideally, these samples could be normalized per seawater volume, but since extraction efficiency is variable and not necessarily reproducible, this simple normalization protocol was deemed deficient. While it was concluded that there was no universal solution to this problem, developing normalizations to "housekeeping" or "core" genes might be better suited for these analyses.

Data and sample archive

An important component of the sectional cruise implementation will be planning for and setting aside material (*i.e.,* DNA, cDNA, and filtered material) in an archive for future research. For example, just ten years ago, scientists would not have imagined that we would be on the brink of implementing such a large sequencing effort in marine

biogeochemistry. This is largely because the technology at the time was too slow and expensive to propose such an effort. In addition to these biotechnological limitations, the computers available at the time were simply not fast enough to assemble the vast amount of information that comes from an environmental metagenome. With this context in mind, workshop participants felt that it was imperative that a sample archive be set up and maintained from a sectional cruise. Not only will there be biological questions that necessitate future analyses, but it is also likely that future technological advances will allow the material from the cruise to be analyzed in new ways that will provide valuable insights on ocean biogeochemistry.

Data synthesis

The synthesis of genomic and geochemical data has never been attempted on this scale before, and at first glance, the veritable avalanche of genomic data alone (e.g., gigabases of sequence data) makes the issue daunting at first. But in principle, such analyses can be broken down into manageable parts, each of which resemble the kinds of studies performed for many years by ecologists and biological oceanographers. We can first ask: How does microbial community composition vary along different geochemical gradients? This includes frequently studied gradients like nutrients, temperature, and light. However, the data we propose to collect will also allow us to test for novel compositional relationships to changes in DOC composition, the concentration of trace metals, and interactions between different trophic levels (e.g., viral abundance or composition). This analysis will be achieved by clustering core phylogenetic markers (e.g., 16S rRNA) into OTUs and then comparing the distributions of those categorized (or binned) sequences with other measured parameters from the same locations by a variety of visualization and statistical techniques. This can be done either with the whole community in mind or with a focus on specific taxa. For example, Martiny et al. (2009) studied the relationship between *Prochlorococcus* diversity and environmental variables by applying canonical correspondence analysis. This analysis revealed that ecotypes at different levels of phylogenetic divergence responded to different environmental factors (Figure 4). Other interactions are likely to be complex, and might best be described and visualized as networks (Fuhrman, 2009). These analyses will reveal the biogeography of marine

microbial communities in relation to many key geochemical variables and can lead to descriptive or predictive models, associating certain collections of organisms with specific conditions or processes (Follows and Dutkiewicz, 2011).
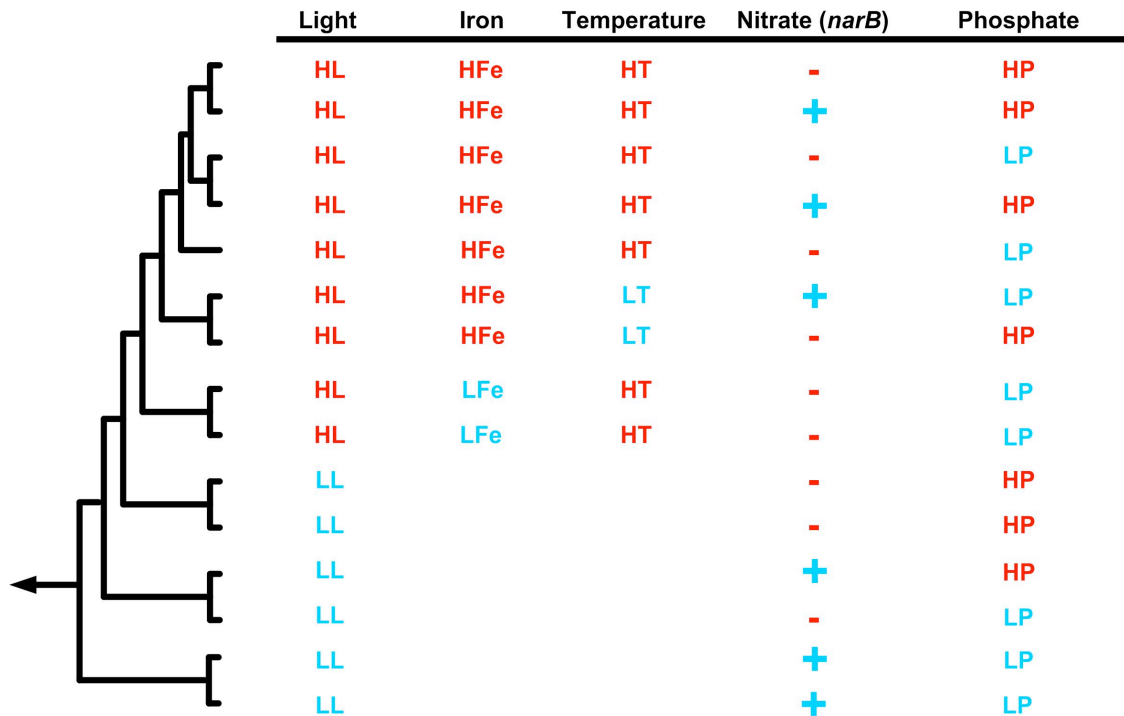
	Light	Iron	Temperature	Nitrate (*narB*)	Phosphate
	HL	HFe	HT	-	HP
	HL	HFe	HT	+	HP
	HL	HFe	HT	-	LP
	HL	HFe	HT	+	HP
	HL	HFe	HT	-	LP
	HL	HFe	LT	+	LP
	HL	HFe	LT	-	HP
	HL	LFe	HT	-	LP
	HL	LFe	HT	-	LP
	LL			-	HP
	LL			-	HP
	LL			+	HP
	LL			-	LP
	LL			+	LP
	LL			+	LP

Figure 4. **Predicted phylogenetic distribution of functional traits in *Prochlorococcus*. Based on laboratory analyses of cultures and field surveys, *Prochlorococcus* can roughly be divided into two groups based on adaptation to different light levels (Moore et al., 2002). The high-light clade can further be divided into two – a high and a low iron clade. This is based on the detection of a previously unrecognized clade in HNLC regions (Rusch et al., 2010). The high-light, high iron clade can be further subdivided into a high and a low temperature group based on the temperature optimum of cultures and only detecting the eMED4 ecotype in surface waters with temperatures below 2^0C (Johnson et al, 2006). Finally, we find high variability in nutrient acquisition gene content among *Prochlorococcus* cultures and field populations (Martiny et al., 2006, 2009), suggesting that adaptation to nutrient availability is associated with fine-scale *Prochlorococcus* clades.**

We then plan to ask: How do changes in microbial diversity influence biogeochemical processes? To address this question, we will analyze how changes in the phylogenetic composition of marine communities relate to changes in the gene repertoire of the community. Here, we will integrate information about phylogenetic composition of the community (e.g., 16S rRNA tags), the identification of functional genes/proteins or networks from meta-genomic, -transcriptomic, or -proteomic data, and finally,

measurements of biogeochemical processes (uptake rates, elemental and structural composition, etc). There is no doubt that this integration will be very challenging and require the development of new ways to analyze data, but this unique dataset has the potential to provide an unprecedented insight into how microbial biodiversity is linked to global ocean biogeochemistry

Concluding thoughts and recommendations

A global sectional survey program would produce a large dataset to characterize the marine microbial community and enable us to characterize its biogeography in a robust way. An important objective of the program is to identify key microbial biomes throughout the world's oceans from surface to seafloor, and attempt to understand how their boundaries are defined. Beyond this core objective, there is a general consensus that virtually every important question in marine microbiology is limited by lack of field data, and this impairs our ability to make accurate predictions about how the microbial community (and thus ecosystems in general) will react to climate change.

A key recommendation is that a stand-alone sectional survey program in microbial biogeography be established. GEOTRACES is a good model in terms of establishing an inclusive process for designing the program and securing funding, as well as establishing an infrastructure for intercalibration and data management. Since GEOTRACES is already established, and the lead time for a new program is long, we recommend a pilot study linking high-resolution metagenomic sampling with high-resolution sampling for GEOTRACES core parameters across pronounced gradients in these parameters.

Workshop participants

Andrew E Allen	J. Craig Venter Institute
Robert F Anderson	Lamont-Doherty Earth Observatory
Elizabeth V. Armbrust	University of Washington/School of Oceanography
Kevin R Arrigo	Stanford University
Katherine Barbeau	Scripps Institution of Oceanography/UC San Diego
Erin M Bertrand	Woods Hole Oceanographic Institution
Nicholas J Bouskill	Lawrence Berkeley National Laboratory
Deborah A Bronk	The College of William & Mary/VIMS
Douglas G Capone	Biological Sciences, USC
Phoebe D Chappell	University of Rhode Island
Maureen L Coleman	Caltech
Allan H. Devol	University of Washington
Chris L Dupont	J. Craig Venter Institute
Clara A Fuchsman	University of Washington
Jed Fuhrman	USC
Robinson W Fulweiler	Boston University
Stephen J Giovannoni	Dept. of Microbiology
Angelos Hannides	University of Cyprus
Cecelia C S Hannides	The Cyprus Institute
Dennis A Hansell	University of Miami
Roberta L Hansman	California Institute of Technology
Karla Heidelberg	University of Southern California
David A Hutchins	University of Southern California
Anitra E. Ingalls	University of Washington
Amal Jayakumar	Princeton University
Bethany D Jenkins	University of Rhode Island
Jon Kaye	Moore Foundation
Kelly A Kryc	GBMF
Elizabeth B Kujawinski	WHOI
Matthew C Long	NCAR

Haiwei Luo	University of Georgia
Elizabeth L Mann	Skidaway Institute of Oceanography
Adrian Marchetti	University of Washington
Roberta L Marinelli	National Science Foundation
Adam C Martiny	University of California, Irvine
Shannon L McCallister	Virginia Commonwealth University
James Moffett	University of Southern California, AHF 204
Jefferson K Moore	University of California, Irvine
Robert Morris	University of Washington
Aurelie C Moya	ARC Centre of Excellence for Coral Reef Studies
Monica V Orellana	Institute for Systems Biology
Elizabeth Padilla	GaTech-Univ of Tennesse
Ann Pearson	Harvard University
Mary Jane Perry	University of Maine
Sophie Richier	National Oceanography Centre
Adam R Rivers	University of Georgia
Julie Robidart	University of CA, Santa Cruz
Gabrielle Rocap	University of Washington
Mak Saito	WHOI
Alyson E Santoro	Woods Hole Oceanographic Institution
Jaclyn K Saunders	University of Washington
Mike Sieracki	Bigelow Laboratory for Ocean Sciences
Jill A Sohm	University of Southern California
Tiantian Tang	Stony Brook University
Benjamin Van Mooy	WHOI
Diana E Varela	University of Victoria
Jacob R Waldbauer	California Institute of Technology
Bess B Ward	Princeton University
Boris Wawrik	University of Oklahoma
Eric A. Webb	University of Southern California, AHF 331
LeAnn P Whitney	University of Rhode Island

References

BEJA, O. and others. 2000. Bacterial rhodopsin: evidence for a new type of phototrophy in the sea. Science **289**: 1902–06.

CHISHOLM, S. W., R. J. OLSON, E. R. ZETTLER, R. GOERICK, J. B. WATERBURY, AND N. A. WELSCHMEYER. 1988. A novel free-living prochlorophyte abundant in the oceanic euphotic zone. Nature **334**: 340–43.

DELONG, E. F., AND D. M. KARL. 2005. Genomic perspectives in microbial oceanography. Nature **437**: 336–42.

FERGUSON, R. L., AND P. RUBLEE. 1976. Contribution of Bacteria to Standing Crop of Coastal Plankton. Limnol Oceanogr **21**: 141–45.

FOLLOWS, M. J., AND S. DUTKIEWICZ. 2011. Modeling Diverse Communities of Marine Microbes. Annu. Rev. Mar. Sci **3**: 427–51.

FRANCIS, C. A., K. J. ROBERTS, J. M. BEMAN, A. E. SANTORO, AND B. B. OAKLEY. 2005. Ubiquity and diversity of ammonia-oxidizing archaea in water columns and sediments of the ocean. P Natl Acad Sci Usa **102**: 14683–88.

FUHRMAN, J. A., AND C. C. OUVERNEY. 1998. Marine microbial diversity studied via 16S rRNA sequences: cloning results from coastal waters and counting of native archaea with fluorescent single cell probes. Aq Ecol **32**: 3–15.

FUHRMAN, J. A. 2009. Microbial community structure and its functional implications. Nature **459**: 193–99.

GIOVANNONI, S., AND U. STINGL. 2005. Molecular diversity and ecology of microbial plankton. Nature **437**: 343–48.

GÓMEZ-CONSARNAU, L. and others. 2007. Light stimulates growth of proteorhodopsin-containing marine Flavobacteria. Nature **445**: 210–13.

JANNASCH, H. W., AND G. E. JONES. 1959. Bacterial Populations in Sea Water as Determined by Different Methods of Enumeration. Limnol Oceanogr **4**: 128–39.

JOHNSON, Z. I., E. R. ZINSER, A. COE, N. P. MCNULTY, E. M. S. WOODWARD, AND S. W. CHISHOLM. 2006. Niche partitioning among *Prochlorococcus* ecotypes along ocean-scale environmental gradients. Science **311**: 1737–40.

KARNER, M. B., E. F. DELONG, AND D. M. KARL. 2001. Archaeal dominance in the mesopelagic zone of the Pacific Ocean. Nature **409**: 507.

KOLBER, Z. S. and others. 2001. Contribution of aerobic photoheterotrophic bacteria to the carbon cycle in the ocean. Science **292**: 2492–5.

KÖNNEKE, M., A. BERNHARD, J. DE LA TORRE, C. WALKER, J. WATERBURY, AND D. STAHL. 2005. Isolation of an autotrophic ammonia-oxidizing marine Archaeon. Nature **437**: 543–46.

KUYPERS, M. M. and others. 2001. Massive expansion of marine archaea during a mid-Cretaceous oceanic anoxic event. Science **293**: 92–95.

KWON, E. Y., F. PRIMEAU, AND J. L. SARMIENTO. 2009. The impact of remineralization depth on the air–sea carbon balance. Nature Geosci **2**: 630–35.

LE QUERE, C. and others. 2005. Ecosystem dynamics based on plankton functional types for global ocean biogeochemistry models. Global Change Biol **11**: 2016–40.

LONGHURST, A. R. 2007. Ecological Geography of the Sea. Academic Press.

MARTINY, A. C., A. P. K. TAI, D. VENEZIANO, F. PRIMEAU, AND S. W. CHISHOLM. 2009. Taxonomic resolution, ecotypes and the biogeography of *Prochlorococcus*. Environ Microbiol **11**: 823–32.

MARTINY, A. C., M. L. COLEMAN, AND S. W. CHISHOLM. 2006. Phosphate acquisition genes in *Prochlorococcus* ecotypes: Evidence for genome-wide adaptation. P Natl Acad Sci Usa **103**: 12552–57.

MEASURES, C.I., LANDING, W.M., BROWN, M.T., BUCK C.S. 2008. High-resolution Al and Fe data from the Atlantic Ocean CLIVAR-CO2 repeat hydrography A16N transect: Extensive linkages between atmospheric dust and upper ocean geochemistry. Global Biogeochem Cycl **22(1).** Article Number: GB1005

MOISANDER, P. H., A. E. MORRISON, B. B. WARD, B. D. JENKINS, AND J. P. ZEHR. 2007. Spatial-temporal variability in diazotroph assemblages in Chesapeake Bay using an oligonucleotide nifH microarray. Environ Microbiol **9**: 1823–35.

MOISANDER, P. H., L. SHIUE, G. F. STEWARD, B. D. JENKINS, B. M. BEBOUT, AND J. P. ZEHR. 2006. Application of a *nifH* oligonucleotide microarray for profiling

diversity of N_2-fixing microorganisms in marine microbial mats. Environ Microbiol **8**: 1721–35.

MOORE, J. K., S. C. DONEY, AND K. LINDSAY. 2004. Upper ocean ecosystem dynamics and iron cycling in a global three-dimensional model. Global Biogeochemical Cycles. Global Biogeochem Cycles **18**: 1–21.

MOORE, J. K., S. C. DONEY, J. A. KLEYPAS, D. M. GLOVER, AND I. Y. FUNG. 2002. An intermediate complexity marine ecosystem model for the global domain. Deep-Sea Res. II **49**:

MOORE, L. R., A. F. POST, G. ROCAP, AND S. W. CHISHOLM. 2002. Utilization of different nitrogen sources by the marine cyanobacteria *Prochlorococcus* and *Synechococcus*. Limnol Oceanogr 47: 989–96**.**

MORAN, M., AND W. L. MILLER. 2007. Resourceful heterotrophs make the most of light in the coastal ocean. Nature Rev Microbiol **5**: 792–800.

PEARSON, A., A. P. MCNICHOL, B. C. BENITEZ-NELSON, J. M. HAYES, AND T. I. EGLINTON. 2001. Origins of lipid biomarkers in Santa Monica Basin surface sediment: A case study using compound-specific Delta C-14 analysis. Geochimica Et Cosmochimica Acta **65**: 3123–37.

RUSCH, D., A. C. MARTINY, C. L. DUPONT, A. HALPERN, AND J. C. VENTER. 2010. Characterization of *Prochlorococcus* clades from iron-depleted oceanic regions. Proc Natl Acad Sci U S A **107**: 16184–89.

RUSCH, D. B. and others. 2007. The Sorcerer II Global Ocean Sampling expedition: Northwest Atlantic through Eastern Tropical Pacific. PLoS Biol. **5**: 398–431-ARTN 077.

SARMIENTO, J. L. and others. 2004. Response of ocean ecosystems to climate warming. Global Biogeochemical Cycles **18**: GB3003.

SCHLEPER, C., G. JURGENS, AND M. JONUSCHEIT. 2005. Genomic studies of uncultivated archaea. Nature Rev Microbiol **3**: 479–88.

TARONCHER-OLDENBURG, G., E. M. GRINER, C. A. FRANCIS, AND B. B. WARD. 2003. Oligonucleotide Microarray for the Study of Functional Gene Diversity in the Nitrogen Cycle in the Environment. Appl Environ Microbiol **69**: 1159–71.

VENTER, J. C. and others. 2004. Environmental genome shotgun sequencing of the Sargasso Sea. Science **304**: 66–74.

WATERBURY, J. B., S. W. WATSON, R. R. L. GUILLARD, AND L. E. BRAND. 1979. Widespread occurrence of a unicellular, marine, planktonic, cyanobacterium. Nature **277**: 293–94.

WEBB, E. A., R. W. JAKUBA, J. W. MOFFETT, AND S. T. DYHRMAN. 2007. Molecular assessment of phosphorus and iron physiology in *Trichodesmium* populations from the western Central and western South Atlantic. Limnol Oceanogr **52**: 2221–32.

WU, L., D. K. THOMPSON, G. LI, R. A. HURT, J. M. TIEDJE, AND J. ZHOU. 2001. Development and Evaluation of Functional Gene Arrays for Detection of Selected Genes in the Environment. Appl. Environ. Microbiol. **67**: 5780–90.

WUCHTER, C. and others. 2006. Archaeal nitrification in the ocean. P Natl Acad Sci Usa **103**: 12317–22.

YODER, J., AND M. KENNELLY. 2005. Global SeaWiFS Chlorophyll Climatology.

YOOSEPH, S. and others. 2007. The Sorcerer II Global Ocean Sampling expedition: Expanding the universe of protein families. PLoS Biol. **5**: 432–66-ARTN 016.