

Techniques for Quantifying Phytoplankton Biodiversity

Zackary I. Johnson¹ and Adam C. Martiny²

¹Marine Laboratory (Nicholas School of the Environment) and Department of Biology, Duke University, Beaufort, North Carolina 28516; email: zij@duke.edu

²Department of Earth System Science and Department of Ecology and Evolutionary Biology, University of California, Irvine, California 92697; email: amartiny@uci.edu

Annu. Rev. Mar. Sci. 2015. 7:299–324

First published online as a Review in Advance on September 17, 2014

The *Annual Review of Marine Science* is online at marine.annualreviews.org

This article's doi:
10.1146/annurev-marine-010814-015902

Copyright © 2015 by Annual Reviews.
All rights reserved

Keywords

phytoplankton diversity, next-generation sequencing, molecular ecology

Abstract

The biodiversity of phytoplankton is a core measurement of the state and activity of marine ecosystems. In the context of historical approaches, we review recent major advances in the technologies that have enabled deeper characterization of the biodiversity of phytoplankton. In particular, high-throughput sequencing of single loci/genes, genomes, and communities (metagenomics) has revealed exceptional phylogenetic and genomic diversity whose breadth is not fully constrained. Other molecular tools—such as fingerprinting, quantitative polymerase chain reaction, and fluorescence in situ hybridization—have provided additional insight into the dynamics of this diversity in the context of environmental variability. Techniques for characterizing the functional diversity of community structure through targeted or untargeted approaches based on RNA or protein have also greatly advanced. A wide range of techniques is now available for characterizing phytoplankton communities, and these tools will continue to advance through ongoing improvements in both technology and data interpretation.

INTRODUCTION

Marine phytoplankton, classically defined as microscopic photosynthetic ($\text{CO}_2 \rightarrow \text{CH}_2\text{O}$) organisms, comprise an exceptionally diverse group of species that includes two domains of life (Blankenship 2002). This genetic diversity translates into fundamental biochemical and physiological differences among clades and species that in turn guide their ecology and their role in ocean biogeochemistry (Falkowski et al. 2008). Phytoplankton perform the vast majority of primary production in the ocean and account for roughly half of the global total production as quantified by oxygen evolution or carbon dioxide uptake (Field et al. 1998). Indeed, it is these dual (and linked) foundational roles of phytoplankton in cycling major elements and funneling the initial biochemically available energy into the oceans that make the topic of phytoplankton biodiversity of fundamental importance for the marine sciences (Falkowski et al. 1998).

The importance of this diversity has been appreciated for centuries by biologists, ecologists, and others, from Ernst Haeckel to G. Evelyn Hutchinson. Thus, the characterization of biodiversity among marine phytoplankton has a long history. Modern molecular techniques have rapidly advanced the ability to characterize and quantify phytoplankton diversity through enhanced precision and throughput, but many of the driving questions remain the same: How many different types (species) of phytoplankton exist in the ocean? What is their evolutionary relationship? How do they coexist? What environmental or biological variables influence their biogeography? How do physiology and behavior vary among species and in response to the environment? And, more broadly, how do the ecology and biogeochemistry differ among these species? More recently, these same measurements and questions have become important components in applied phycology, which can leverage this biodiversity to achieve a specific outcome (biofuels from algae, harmful algae bloom identification/prediction, etc.) (Lang et al. 2011, Parsons et al. 1999).

Operationally, phytoplankton biodiversity can be subdivided into three different but interrelated components: genomic, phylogenetic, and physiological. Genomic diversity broadly encompasses the instruction set or range of capabilities that a given phytoplankton has at its disposal, including its entire gene complement (genome) as well as any potential epigenetic properties. Phylogenetic (or taxonomic) diversity stems from genomic diversity but is focused more on the comparative evolutionary relationships among different phytoplankton species. Functional diversity broadly characterizes the biochemical, physiological, and behavioral responses to any range of environmental variables, from temperature and light to viral susceptibility to biochemical composition (among many others). In practice, these overlapping components of biodiversity provide mileposts along the way to the broader goal of characterizing phytoplankton community biodiversity in order to understand what is out there, how are they related, and what are they doing.

HISTORICAL APPROACHES

Although phytoplankton en masse (e.g., blooms) have long been observed, the development of microscopy by Robert Hooke and Antonie van Leeuwenhoek enabled the first detailed characterization of phytoplankton biodiversity by direct observation. Their fledgling microscopes had limited optical resolution but nevertheless provided a more detailed view of morphology, which in turn allowed an initial characterization of plankton's taxonomic diversity. Perhaps most famous are the detailed drawings made by Haeckel [2005 (1862)], which provide exceptional resolution of the distinct structures of silicified organisms such as diatoms and radiolarians. This direct morphological characterization still represents an important technique in characterizing phytoplankton biodiversity, especially for larger cells that have distinguishable morphologies (**Figure 1c**).

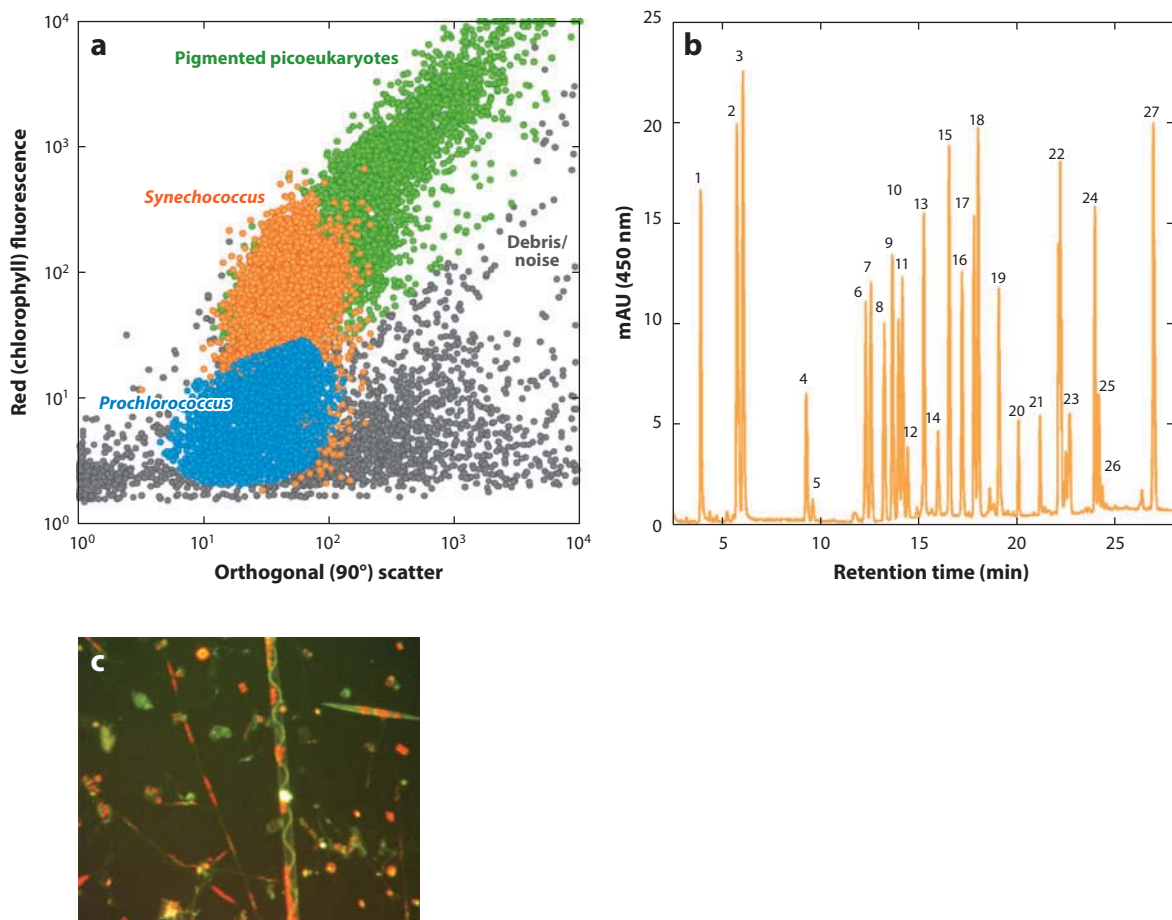


Figure 1

Examples of optical characterization of phytoplankton diversity. (a) Flow cytogram of a mixed population of picoplankton. Each colored dot represents an individual cell of an identified taxon. (b) Pigment separation and quantification using high-performance liquid chromatography. Numbered peaks represent distinguishable pigments. Data are from Bidigare et al. (2005). (c) Microscope image of a sample dominated by eukaryotic phytoplankton, demonstrating some of the morphological diversity useful in assigning taxonomy. Image courtesy of Karen Selph.

Improvements in light microscopes as well as the development of the electron microscope helped to form many of the early descriptions of marine phytoplankton. These approaches, which continue today, have been notably successful for diatoms, prymnesiophytes, and other taxa with ornate morphologies. Recent advances in superresolution fluorescence microscopy will no doubt further extend this approach, and in particular will aid studies of physiological and behavioral biodiversity because of their ability to directly resolve processes at the cellular level (Westphal et al. 2008).

In parallel, other early optical characterizations of algae relied on basic colorimetric approaches. All phytoplankton contain chlorophyll *a* (either in monovinyl or divinyl form) as their core pigment for harvesting light. As such, it is generally not useful as a metric for biodiversity, but many accessory pigments such as carotenoids or other types of chlorophyll (e.g., chlorophyll *b* or *c*) are taxon specific, and many lineages have unique optical spectra. Early work classified algae into four

Table 1 Taxon-diagnostic pigments as quantified by high-performance liquid chromatography (compiled from Bidigare et al. 2005 and Mackey et al. 1996)

Pigment	Associated taxa
19'-Butanoyloxyfucoxanthin	Chrysophytes and haptophytes
19'-Hexanoyloxyfucoxanthin	Chrysophytes and haptophytes
Neoxanthin	Prasinophytes, euglenophytes, and chlorophytes
Chlorophyll <i>b</i>	Prasinophytes, euglenophytes, <i>Prochlorococcus</i> , and chlorophytes
Violaxanthin	Prasinophytes and chlorophytes
Diadinoxanthin	Dinoflagellates, chrysophytes, haptophytes, euglenophytes, and diatoms
Lutein	Prasinophytes and cyanobacteria
Zeaxanthin	Prasinophytes, cyanobacteria, and chlorophytes
Divinyl chlorophyll <i>a</i>	<i>Prochlorococcus</i>

groups—red, green, brown, and diatomaceous algae—based solely on their appearance. Although not focused on phytoplankton (i.e., single-celled organisms) per se, this approach formed the foundation for pigment-based analyses that are still performed today.

For example, chromatography-based approaches and, specifically, modern high-performance liquid chromatography (HPLC) can simultaneously determine the concentrations (and therefore approximate abundance) of a wide range of carotenoids and chlorophylls (Bidigare et al. 2005) (**Figure 1b**). The abundance of these diagnostic accessory pigments relative to the universal chlorophyll *a* pigment can be used to quantify the abundance of major taxa in a sample (Mackey et al. 1996) (**Table 1**). These markers are generally not unique to specific groups, and precise identification is therefore challenging. Furthermore, pigment ratios are strongly influenced by light, nutrient availability, and other environmental variability. Thus, these ratios can potentially provide information on the physiological diversity of phytoplankton, but they have only limited applicability to the precise quantification of specific taxa. Nevertheless, because of their relatively high throughput and low cost, chromatography-based approaches are useful for large numbers of samples to obtain a broad taxonomic characterization of a given community.

Early use of flow cytometry also enabled the optical characterization of phytoplankton communities (Yentsch et al. 1983). Single cells are passed through a focused laser beam (or beams) using a laminar flow stream. Scattered laser light, which is proportional to cell size and refractive index, is then used to quantify and characterize cells. In addition, fluorescence can be diagnostic for some major pigments groups (e.g., chlorophyll and phycoerythrin). The combined approximate size/refractive index and pigment composition can be used to identify several major taxa of phytoplankton, including *Prochlorococcus*, *Synechococcus*, and pennate diatoms, and can enumerate other apparent populations even if they are not taxonomically resolved (Olson et al. 1985) (**Figure 1a**). Flow cytometry typically works best on smaller cells (less than ~20 μm in size) that are approximately spherical, and it continues to be the most robust technique for counting major picoplankton populations (Lomas et al. 2011, Shapiro 2003). For larger cells (greater than ~10 μm in size) from morphologically distinct populations, imaging cytometry is preferred. This method marries microscope-like optical (i.e., visual) characterization with a flow cytometry-like sample stream, thus providing much higher throughput than standard microscopic analysis (See et al. 2005, Sieracki et al. 1998). Pictures (or other optical information like fluorescence) are obtained from each cell as it crosses the imaging point, and the cells are later identified either manually or automatically based on training data sets. The in situ video plankton recorder uses a similar approach to image cells as they pass through the interrogation point, with the additional advantage that the instrument is towed behind a ship (Benfield et al. 2007).

As digital cameras and other optical sensors continue to improve in sensitivity, resolution, speed, size, and durability, flow cytometry and imaging cytometry will be able to provide even more optical information on phytoplankton populations (and their component single cells). Instruments have also been developed for extended in situ deployment, providing exceptional temporal resolution (Dubelaar et al. 1999). The storage of raw data and automated processing of signals (including images) remain major challenges, but work in this area has progressed, in part because of efforts to overcome similar challenges in other fields that use imaging technologies (security, medicine, etc.).

MOLECULAR APPROACHES

Optical approaches dominated phytoplankton biodiversity studies until approximately the turn of the millennium; since then, there have been many technological advances in the field. Many of these advances have focused on molecular characterization, enabling increased throughput and/or deeper characterization of samples. For example, DNA sequencing was often previously limited to single loci and low sample throughput, but common approaches can now generate gigabases of data at low cost. Similar advances in mass spectroscopy have advanced the identification and quantification of proteins as well as tracer-based isotopic analyses for the characterization of physiological diversity. In addition to these and other methodological advances, exponential increases in computational power, data storage, and data processing routines have been critical. They have allowed the acquisition, analysis, and ultimately interpretation of the often vast amounts of data generated by the new technologies. In the sections that follow, we summarize some of the major technological advances and then discuss how these tools can be used to characterize biodiversity in each of its forms (**Table 2**). Finally, we discuss emerging techniques, including how different techniques are being merged to advance our understanding of biodiversity.

DNA Sequencing

Of the modern techniques available, perhaps the most well developed and extensively used in the study of biodiversity is DNA sequencing. DNA sequencing has evolved considerably from the initial and tedious slab gel approach and now includes massively parallel approaches that enable many downstream applications. Marine scientists use several forms of DNA sequencing (including RNA sequencing via reverse transcription) depending on the application, budget, and other requirements. Sanger (or chain termination) sequencing is still considered the gold standard because it has a moderate read length [~ 800 base pairs (bp)] and the data are of high quality, but it also carries the disadvantages of low throughput and high cost. It is now used primarily for single-locus sequencing of cultures and other low-diversity applications (e.g., clone libraries) where a small number of high-quality sequences are needed.

As an alternative to Sanger sequencing, so-called next-generation sequencing can utilize any of five different technologies (Mardis 2013). Pyrosequencing (454 Life Sciences/Roche Diagnostics) provides moderate read lengths (up to ~ 700 bp), but it is relatively expensive and provides fewer reads (~ 1 million) per run compared with other methods. Because it was the first widely available next-generation technique, many pioneering phylogenetic and genomic diversity studies used this approach. Sequencing by ligation (SOLiD platform, Thermo Fisher Scientific/Life Technologies) can provide more than 1 billion high-quality, shorter reads ($50 + 50$ bp) at a low cost per base, but it is generally slower than other methods. To date, it has not been used extensively by the marine science community. Sequencing by synthesis (Illumina/Solexa) generates up to 3 billion bases of low to moderate read length ($50\text{--}300$ bp) with moderate accuracy at a very low cost. Because of the

Table 2 Phytoplankton biodiversity techniques: applications and properties

Technique	Applications	(Phylo)genetic resolution	Throughput	Dynamic range	Limitations	Example reference(s)
Phylogenetic diversity						
Microscopy (light, fluorescence, electron)	Large, distinct cells; few samples; direct counts	Morphologically distinct	Low	Limited	Large cells	Malfatti & Azam 2009, Thompson et al. 2012
Pigments (HPLC)	Many samples	Coarse	Medium	High	Influenced by physiology	Bidigare et al. 2005, Mackey et al. 1996
Flow cytometry	Many samples; direct counts	Coarse	Medium	High	Small cells (<~20 μm)	Olson et al. 1985
Imaging cytometry, video plankton recorder	Many samples; direct counts	Coarse	Medium	High	Large cells (>~10 μm)	See et al. 2005, Sieracki et al. 1998
Marker genes: ARISA, DGGE, TRFLP	Many samples; approximate counts	Medium	High	Medium	DNA/PCR based	Brown et al. 2005, Moeseneder et al. 1999, Zeidner & Beja 2004
Marker genes: clone library	Few samples	Medium (depending on the product and length)	Low	Medium	DNA/PCR based	Moon-van der Staay et al. 2001, Pichard et al. 1997
Marker genes: amplicon library	Many samples; approximate counts	Medium (depending on the product and length)	High	Medium	DNA/PCR based; significant data analysis	Hunt et al. 2013, Sogin et al. 2006
Marker genes: qPCR	Many samples; direct counts; specific targets	Variable (depending on the PCR primers)	High for samples, low for groups	High	DNA/PCR based	Johnson et al. 2006
Marker genes: hybridization	Many samples; approximate/direct counts; specific targets	Variable resolution (depending on the oligonucleotides)	High for samples, low for groups	High	Predefined targets	Amann & Fuchs 2008
Multiple marker genes	MLST	High	Low	NA	Requires culture or a large proportion of the genome sequence	Maiden et al. 1998

Genomic diversity							
Genomics	Domesticated population (or sorted cells)	High	Low	NA	Domestication of cells (or missing parts of genome if sorted cells); significant data analysis	Armbrust et al. 2004, Rocap et al. 2003	
Metagenomics	Few samples	High	Low	NA	Significant data analysis	DeLong et al. 2006, Venter et al. 2004	
Physiological diversity							
HPLC pigment labeling	Few samples	Coarse	Low	High	Incubation based	Goericke & Welschmeyer 1993	
RT-qPCR (single target)	Many samples; direct quantification; known targets	Variable (depending on the PCR primers)	High for samples, low for groups	High	RNA/PCR based; intermediate step to physiology	Church et al. 2005, Paul et al. 1999	
RT-amplicon library	Many samples; approximate quantification; known targets	Variable (depending on the PCR primers)	High	Medium	RNA/PCR based; intermediate step to physiology; significant data analysis	Campbell et al. 2011, Hunt et al. 2013	
Microarray	Few samples; quantification; many/known targets	High	High for targets, low for samples	High	Intermediate step to physiology	Martiny et al. 2006, Zinser et al. 2009	
Transcriptomics	Few samples; quantification	High	High for targets, low for samples	Medium	Intermediate step to physiology	Allen et al. 2008	
Metatranscriptomics	Few samples; approximate quantification; unknown targets	High	High for targets, low for samples	Low	Significant data analysis; intermediate step to physiology	Poretzky et al. 2009	
Proteomics (single or multiple proteins)	Few samples; approximate quantification; known targets	Medium	High for targets, low for samples	Low	Significant data analysis (if multiple proteins); intermediate step to physiology	Dyhrman et al. 2012, Hockin et al. 2011	
Metaproteomics	Few samples; approximate quantification; known targets	Medium	High for targets, low for samples	Low	Significant data analysis; intermediate step to physiology	Morris et al. 2010	

(Continued)

Table 2 (Continued)

Technique	Applications	(Phylo)genetic resolution	Throughput	Dynamic range	Limitations	Example reference(s)
Physiology (growth, behavior, etc.)	Intermediate samples	Low	Low	Variable	Domestication of cells	Falkowski & Owens 1980, Franks 1992
Biochemistry (elemental composition, fatty acids)	Variable samples	Low	Low	Variable	Variable; often requires domestication of cells	Bertilsson et al. 2003, Haldal et al. 2003, Van Mooy et al. 2006
Tracer uptake (e.g., BrdU, isotopes)	Many samples; quantification	Low	Low	Variable	Incubation based	Casey et al. 2007
Metabolites	Few samples; quantification	Low	Low	Medium	Incubation based; significant data analysis	Osanai et al. 2014
Other						
Cell sorting (flow cytometry, microfluidics)	Any downstream single-cell-based (or population-based) approach	Variable	Low	Low	Ability to optically defined cells (with or without stained probes)	Malmstrom et al. 2013
Single-cell DNA amplification (e.g., MDA)	Any downstream DNA-based approach	Variable	Low	Low	Partial genome coverage	Zhang et al. 2006

Abbreviations: ARISA, automated ribosomal intergenic spacer analysis; BrdU, bromodeoxyridine; DGGE, denaturing gradient gel electrophoresis; HPLC, high-performance liquid chromatography; MDA, multiple displacement amplification; MLST, multilocus sequence typing; NA, not applicable; PCR, polymerase chain reaction; qPCR, quantitative polymerase chain reaction; RT, reverse transcription; TRFLP, terminal restriction fragment length polymorphism.

high throughput and low cost, many microbial diversity studies, including those for phytoplankton, are moving to this technology. Semiconductor sequencing (Ion Torrent platform, Thermo Fisher Scientific/Life Technologies) rapidly (~hours) generates up to 80 million reads of low to moderate length (100–400 bp) with moderate accuracy at an intermediate cost. To date, it has not been used extensively by the microbiological or marine science community. Finally, single-molecule sequencing (Pacific Biosciences) rapidly (hours) generates up to 50,000 long reads (~5,000–30,000 bp) of low quality at moderate expense. To date, this technology has not been used by the marine science community, but it holds promise for several applications that can leverage the longer reads.

Of the six major techniques available to the marine science community, Sanger sequencing has historically been the most used. However, its application has rapidly declined as new technologies have emerged. For example, pyrosequencing has been used extensively in the past 10 years because it was the first next-generation technique available and greatly reduced the per-base cost, and sequencing by synthesis is now rapidly becoming the most popular approach because of its even lower per-base cost. In this rapidly advancing technological landscape, other approaches (e.g., single-molecule, semiconductor, and other nascent technologies) may ultimately become more common as capabilities grow and costs decline.

Phylogenetic (Taxonomic) Diversity

Although many optical techniques are still used today, molecular approaches have rapidly replaced them as the preferred method for characterizing phytoplankton biodiversity. Nowhere is this more apparent than in the analysis of phylogenetic/taxonomic diversity, where a suite of approaches based on marker genes (single loci) are available. Many of these techniques rely on DNA sequence approaches in which a locus of interest (often a gene or gene fragment) is amplified or hybridized for quantification or other characterization. These approaches require a known target (DNA sequence) for amplification or targeting and thus can be limited to assessing communities where this target is known and conserved. For example, conserved regions in the photosystem II D1 gene (*psbA*) or the small-subunit RNA gene (16S or 18S) are used to amplify gene fragments for a community of interest, and these fragments are then quantified or otherwise characterized (Hunt et al. 2013, Sogin et al. 2006, Zeidner et al. 2003).

In principle, this approach is straightforward, but in practice some care must be taken to avoid bias. For example, designing oligonucleotide primers for target sequences that are specific enough to amplify only target regions of the desired group (diatoms, all phytoplankton, cyanobacteria, *Synechococcus* clade IV, etc.) but general enough to capture all the diversity within the target group can be very challenging. Direct sequencing makes it straightforward to determine whether the approach is yielding members outside the group of interest; however, assessing whether the approach is missing members is often difficult. Nevertheless, many high-quality target sequences have been identified and robust primers or probes developed (e.g., for 16S rRNA) (Hunt et al. 2006, Ludwig et al. 2004). A second potential bias relates to the need for PCR amplification (Chandler et al. 1997, Polz & Cavanaugh 1998). However, PCR errors can be reduced using a variety of approaches, including minimizing the level of amplification and using high-fidelity reagents and methods (Thompson et al. 2002). A third potential complication is that organisms often contain more than one copy of a given gene, and often the sequence (or length) of these copies is different (Farrelly et al. 1995), thus complicating interpretation in mixed communities. With appropriate care in primer/probe design and amplification, many downstream applications—including fingerprinting (see below) and sequencing (cloning or amplicon libraries)—have dramatically expanded the tool kit available for studying the diversity of phytoplankton.

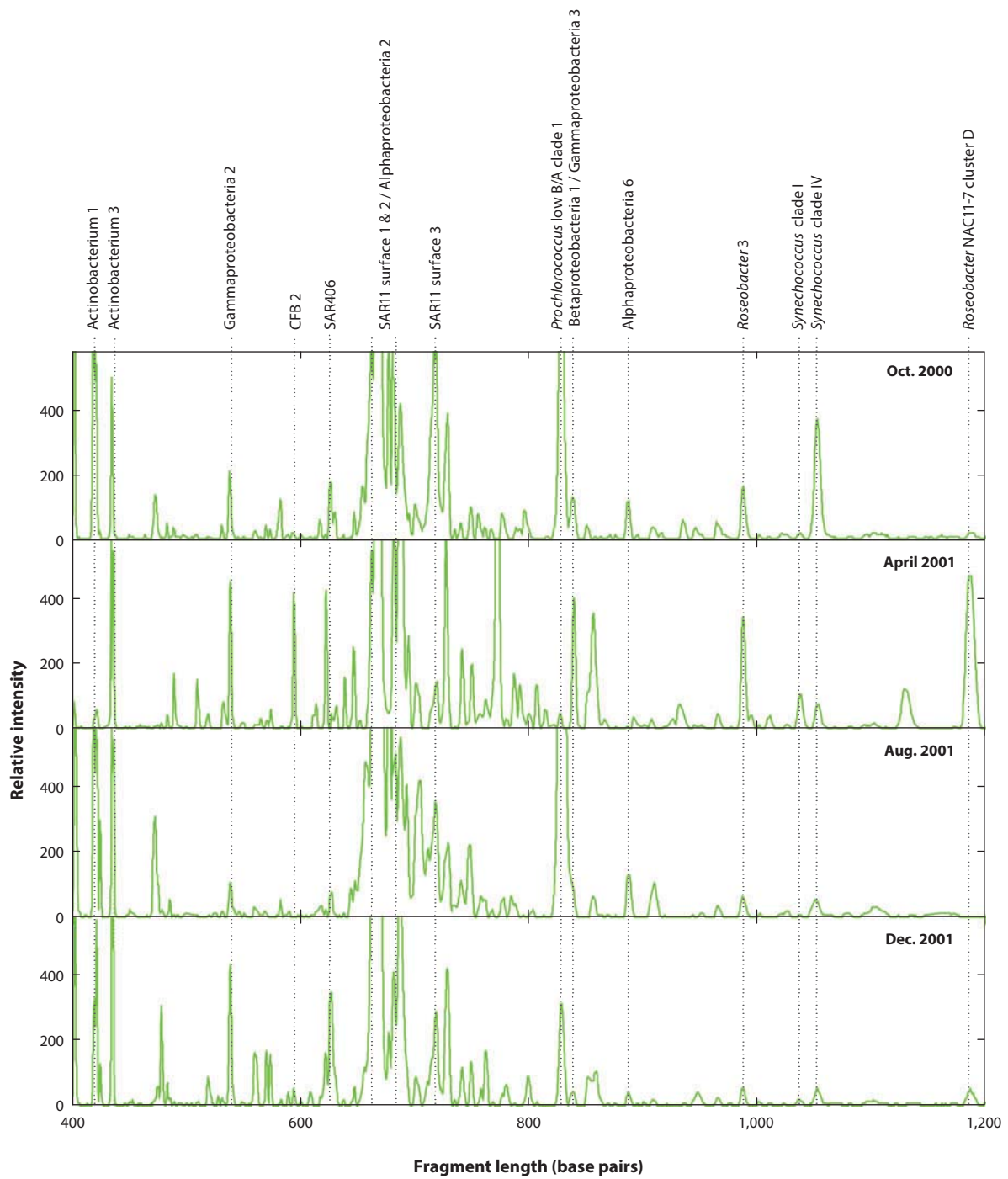
Marker genes: fingerprinting (ARISA, DGGE, and TRFLP). Using marker genes such as *psbA*, *rbcL* (the RuBisCO large subunit gene), or the small-subunit rRNA gene, fingerprinting techniques leverage DNA sequence diversity without requiring direct sequencing of those fragments. The three most commonly used approaches are automated ribosomal intergenic spacer analysis (ARISA), denaturing gradient gel electrophoresis (DGGE), and terminal restriction fragment length polymorphism (TRFLP), all of which rely on amplifying target fragments that are subsequently characterized. ARISA uses the variability in the length of the fragment between the 16S and 23S rRNA genes as an index of taxonomy (**Figure 2**). DGGE uses the migration distance along a denaturing gradient gel as an index of taxonomy. TRFLP uses terminal (either 5' or 3') fragment length after one or more restriction digests as an index of taxonomy.

These techniques are limited in their phylogenetic resolution of sequence, and their dynamic ranges for different populations within a mixed community are poor. However, they are relatively straightforward and economical to use and therefore have been particularly powerful for comparing samples. For example, time-series analysis of high-resolution ARISA signatures has provided substantial insight into microbial populations (Brown et al. 2005, Fuhrman et al. 2006). Similarly, investigators have used DGGE and restriction mapping of phytoplankton marker genes (e.g., *psbA*) to identify novel groups of phytoplankton and their variability in the environment (Scanlan et al. 1996, Zeidner & Beja 2004).

Marker genes: sequencing (clones, amplicons). Because of its increased information content, direct sequencing of gene fragments remains the most robust approach for characterizing taxonomic diversity when using a single locus. Until recently, most approaches relied on cloning fragments into a vector that was transformed (usually into *E. coli*) and subsequently sequenced. The advantages of this approach include high sequence quality and the ability to resequence to extend or verify ambiguous sequence data. One potential difficulty is that gene fragments (or whole genes) are transformed into living cells; although these genes or gene sequences are generally robust, some are not viable in the host and so are not recovered in the final analysis, thus introducing bias. Nevertheless, this cloning and sequencing approach greatly advanced the field because it identified a high level of taxonomic diversity that had not been quantified by optical approaches. For example, initial clone libraries identified unexpected diversity in the 18S rRNA gene, including new clades and taxa (Moon-van der Staay et al. 2001). Other foundational studies identified major taxonomic groups of marine microbes that previously were unrecognized (Giovannoni et al. 1990). However, because of the processing required and the cost, this approach does not scale well for many samples and sequences. It also has relatively poor dynamic range within a sample, and thus is currently used mainly for pilot or other small-scale studies.

Amplicon libraries sequenced using next-generation techniques have greatly expanded single-locus sequencing. Amplicon libraries avoid the bias introduced by cloning and can be sequenced to far greater depths than cloning-based approaches. For example, a large clone library might include several hundred sequences, whereas an amplicon library can now exceed tens of thousands or more (Hunt et al. 2013, Sogin et al. 2006). Although amplicon fragments are somewhat reduced in sequence length (often called tags) and therefore provide reduced phylogenetic resolution, well-selected targets still provide substantial information on diversity. Early use of this approach

osomal intergenic spacer analysis (ARISA) profiles of phytoplankton biodiversity for four time periods. Each peak represents a phytoplankton taxon (or mixture of taxa), and the peak height (or area) is roughly proportional to abundance. Data are from Johnson et al. (2005).



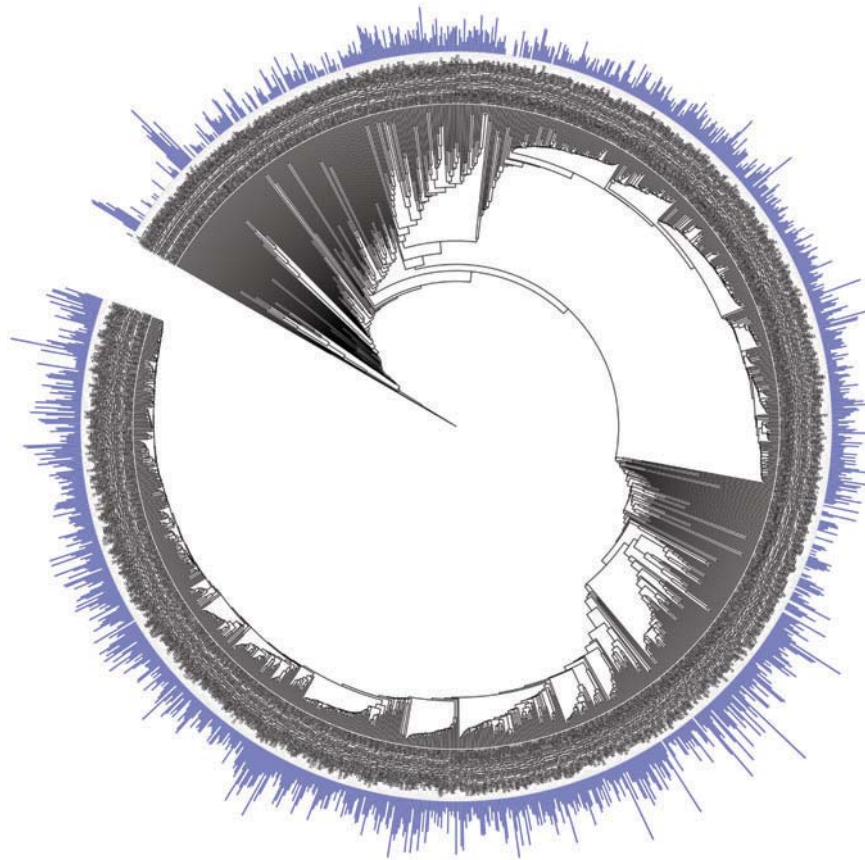


Figure 3

Amplicon library of cyanobacteria internal transcribed spacer sequences, represented as a phylogenetic tree (*inner black lines*) with relative abundances (*outer blue bars*).

focused on using pyrosequencing or small fragments (<100 bp); more recently, sequencing by synthesis, which generates many more and longer reads, has become commonplace (e.g., Caporaso et al. 2011). Depending on the platform and technology, both of which are rapidly changing, these approaches can currently yield up to several million sequences per run with lengths of up to 300 bp (or even more). Because of the vast number of sequences generated, samples are often combined (or bar coded) into a single run and later computationally differentiated by identifying the added bar-code sequences. Broadly, these approaches have shown that microbes and phytoplankton are exceptionally taxonomically diverse, with significant microphylogenetic variability (e.g., **Figure 3**) (Cordero & Polz 2014).

Data analysis. Many techniques for assessing the biodiversity of phytoplankton use DNA sequencing-based approaches. The exponential increases in sequencing capabilities (and the decreased costs) have enabled the production of vast quantities of data. Although this presents great opportunities for characterizing biodiversity, it also presents at least three challenges for data analysis. First, as with all forms of data, DNA sequencing is not error free. Both random and systematic errors can dramatically affect downstream results, and so proper care must be taken

to minimize these errors. These issues are often specific to individual sequencing platforms and should be addressed in this context (Bragg et al. 2013). Second, quality-controlled data are typically processed to identify groups of phytoplankton, either de novo through the creation of operational taxonomic units based on sequence similarity (e.g., 97% similarity) or by binning sequences into known taxa based on a priori information (Preheim et al. 2013, Sogin et al. 2006). This grouping serves the dual purpose of identifying taxa and quantifying their abundance, but it also greatly reduces the size of the data set to representative sequences, thus facilitating downstream analyses. Third, the phylogenies of these representative sequences are typically compared, either by using established relationships (i.e., mapped onto a phylogenetic tree) or by creating de novo relationships (Hall 2011). From these processed (and reduced) sequence data, investigators can perform other biodiversity analyses, such as visualization and the calculation of diversity indices (Escalas et al. 2013). Because of the widespread use of amplicon libraries, many analysis pipelines, such as mothur and QIIME (Quantitative Insights into Microbial Ecology) (Caporaso et al. 2010, Schloss et al. 2009), have been developed that combine the initial processing, operational taxonomic unit grouping, and downstream analyses (e.g., clustering) for single-locus sequences. Other programs used in general ecological analysis [e.g., Primer and R (including the vegan package)] as well as phylogenetic programs [e.g., PHYLIP (Phylogeny Inference Package)] can also be useful for downstream biodiversity analysis.

Marker genes: quantification (qPCR, hybridization). Although fingerprinting and sequencing-based approaches using amplicons provide significant information about the phylogenetic composition of a sample, they can be limited in their ability to precisely quantify the abundance of specific lineages. Quantitative polymerase chain reaction (qPCR) and hybridization-based approaches (at either the cellular or extracted DNA level) can more precisely quantify specific components of taxonomic diversity (Amann & Fuchs 2008, Zinser et al. 2006). These techniques rely on the quantification of known genetic targets (i.e., marker gene sequences as PCR primers or oligonucleotide probes) and therefore are limited to characterizing known diversity. Furthermore, because they require separately assaying each target, these techniques are most commonly used to quantify from one to tens of known groups. However, these approaches can have great dynamic range and, when optimized, can detect from as few as one cell per assay up to the most abundant population in a mixed group. When hybridization is used in conjunction with microscopy [e.g., fluorescence in situ hybridization (FISH)], potential interactions, associations, or localization information can also be available.

Because of these properties, qPCR and hybridization-based approaches have been most successfully employed when quantifying known groups over many samples in which the dynamic range of the target covers several orders of magnitude. For example, Johnson et al. (2006) quantified six genetic clades of *Prochlorococcus* over the Atlantic using qPCR and revealed biogeographic niche partitioning, which was subsequently related to environmental variables. When compared with independent methods for tracking *Prochlorococcus* communities, the sum of the qPCR populations generally agreed well with the total *Prochlorococcus* cell count, but qPCR also identified environments where unquantified diversity remained (i.e., where the qPCR count was significantly less than the total number of cells in the *Prochlorococcus* community).

Genomic Diversity

Phylogenetic diversity provides insight into the range of types (species, clades, etc.) of phytoplankton present in a given sample, and to a limited degree this information can be translated into major functional potential based on past knowledge (e.g., diatoms assimilate silicon). However,

genome-based approaches provide a detailed road map of the genetic biodiversity based on the complement of genes and associated functions. Over the past two decades, a range of techniques have been developed targeting either single organisms or communities, and these have provided major insights into phytoplankton biodiversity.

Genomics. Whole-genome sequencing typically uses a combination of molecular approaches to determine the complete genome sequence of the organism of interest. In most cases, shotgun sequencing is used: Genomic DNA is randomly broken into smaller fragments, which are then sequenced and computationally reassembled to yield the original whole-genome sequence. This approach has been used mainly on model members of major groups of phytoplankton, including diatoms, cyanobacteria, and prymnesiophytes (Armbrust et al. 2004, Bowler et al. 2008, Read et al. 2013, Rocap et al. 2003). The power of this approach lies in two major areas: (a) using the similarity of the DNA sequence to known sequences (i.e., from databases such as GenBank and CyanoBase) to infer the functions of genes and ultimately the range of functions of an entire genome/organism, and (b) using comparative genomics to infer the similarities and differences between two or more different genomes/organisms. For example, early genomic sequencing of cyanobacteria (two *Prochlorococcus* strains and one *Synechococcus* strain) characterized the shared genomic component (or so-called core genome) and highlighted genomic and functional similarities among marine cyanobacteria (Palenik et al. 2003, Rocap et al. 2003). These comparisons also yielded critical differences in the genome content among different strains: Many genes were of known function and therefore could be used to infer differences in ecology and biogeochemistry among cyanobacteria.

As additional strains representing different clades have been sequenced, both the core genome and the unique functional potential found in each clade or organism have been refined and additional insight gained into their ecology and biogeochemistry (Kettler et al. 2007, Scanlan et al. 2009). Eukaryotic phytoplankton typically have much larger genomes that can range from tens of megabases to more than 1,000 megabases (thus potentially exceeding the size of the human genome). This can present significant technical (and financial) challenges and has thus limited the availability of genome sequences from eukaryotes. Nevertheless, the initial genome sequences have provided great insight into the vast metabolic complexity of eukaryotic phytoplankton (Bowler et al. 2008, Palenik et al. 2007). Genomic studies can also expand our understanding of the phylogenetic diversity and evolutionary history of phytoplankton (Falkowski et al. 2004). In particular, plastid sequences from eukaryotes as well as specific regions of the chromosome can be used to probe the timing and sequence of speciation (Worden et al. 2009).

As sequencing prices continue to drop and technology becomes more advanced, genomic characterization will in all likelihood become more common and be applied to an increasing number of strains. If sufficient numbers of members of the population are assessed, this approach could potentially be used in population genetic studies. It is notable that all genomes contain a substantial fraction of genes whose functions remain unknown. These genes, although phylogenetically informative, are not yet useful in characterizing functionality, but they do present genetic targets that could be further explored to quantify differences in biodiversity among phytoplankton species and environments (Karl 2007).

Metagenomics. Because of biomass requirements, genomics has typically been restricted to organisms in culture. However, related metagenomic analyses have been used in mixed populations, characterizing the genetic diversity at the community level. Metagenomic techniques for identifying the genomic diversity within microbial communities fall into two main categories: probes and sequencing. The first category is based on the development of genetic probes that are subsequently

attached to a surface (typically nylon, glass, or silicon). This array-type technique was originally developed to examine patterns of gene expression in individual organisms, whereby probes were designed to target each gene (see below). Based on this concept, arrays have now been designed to target specific genes in mixed communities. These array-based metagenomic techniques were first developed for soil microbial communities (e.g., GeoChip) (He et al. 2007) but were then later designed for marine communities, including phytoplankton (Ottesen et al. 2011, Rich et al. 2008, Shilova et al. 2014).

One of the challenges of array-based techniques is the need for a priori design of probes targeting all the genetic variants of each gene as well as possible differences in binding affinity and associated signal strength. We have only a limited understanding of the microbial diversity in many regions of the ocean, which leads to uncertainty in how well probes will match the genomic variability in a sample. Compared with other techniques, however, arrays are typically less expensive, take less time for analysis, and require less computational power. They are therefore particularly useful for repeated analyses of a genomically well-characterized community or region (e.g., through a time series). Furthermore, an array can be mounted on a buoy or autonomous vehicle to allow for at least semicontinuous temporal analyses (Robidart et al. 2012, Scholin et al. 2009), thus facilitating a genomics-based approach to environmental monitoring as well as identifying short-term changes.

Metagenomic sequencing is another powerful technique for assessing phytoplankton communities and has transformed our understanding of marine diversity. The basic principle has been well established, involving shearing DNA into pieces, sequencing it, and performing an extensive data analysis (DeLong et al. 2006, Venter et al. 2004). This approach provides a reasonably unbiased view of biodiversity and thus allows for the detection of unknown lineages or functions. There are multiple variants of metagenomic sequencing. The first generation of metagenomics techniques cloned pieces of DNA into vectors (plasmid, fosmid, or bacterial artificial chromosome), then amplified and sorted these vectors by transforming them into *E. coli* and isolating individual colonies (DeLong et al. 2006, Venter et al. 2004). Each end of the DNA insert was then sequenced. The second generation of metagenomics has utilized next-generation high-throughput sequencing, which removes the need for cloning and its associated bias (see above). This technique has greatly reduced the cost of metagenomics and generally increased the coverage of organisms captured in a sample. However, the shorter sequence length has also made the downstream analysis more complex and introduced more uncertainty into the taxonomic and gene function assignments.

A variant of metagenomics utilizes flow cytometry-based cell sorting to target specific populations (Batmalle et al. 2014, Palenik et al. 2009, Zehr et al. 2008). However, owing to practical limitations in the number of cells sorted, this approach leads to vanishingly small amounts of the DNA. Thus, the sequencing of sorted populations requires DNA amplification, which can be achieved by either multiple displacement amplification (MDA) or a transposon-based technique (e.g., Nextera) (Batmalle et al. 2014). It has been well established that MDA leads to an extensive bias whereby the representations of some genomic regions are highly skewed. This is not a large issue if the main goal is to discover new diversity, but it can become important if the aim is to compare diversity between populations.

Independent of the details of amplification, the power of these targeted approaches lies in their ability to identify the genomic variation in populations of low overall density (e.g., Zehr et al. 2008) and to associate unknown genes with specific taxa. Flow cytometry-based sorting combined with genome amplification can also be used to target individual cells (Chitsaz et al. 2011, Woyke et al. 2009). Genome sequencing of single cells was first demonstrated in *Prochlorococcus* (Zhang et al. 2006) and is a powerful approach for gaining insights into the combination of genes in a cell and, more broadly, in a lineage. The limitation of the technique is linked to the MDA amplification bias

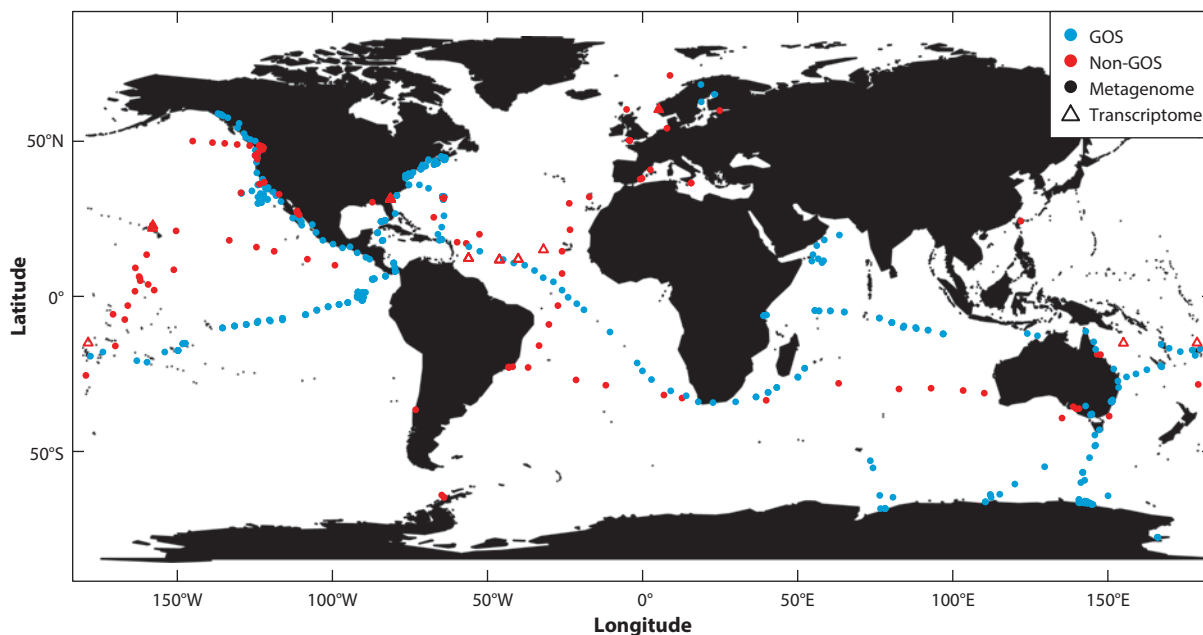


Figure 4

Global distribution of ocean metagenomic and metatranscriptomic samples. The map includes all samples from the Global Ocean Sampling (GOS) survey (Rusch et al. 2007) as well as samples deposited in MG-RAST (Meyer et al. 2008).

(Rodrigue et al. 2009), which so far has impeded the sequencing of a complete genome. However, partial genome sequences can be powerful for linking taxonomic and functional information, and they have been used to explore the genomic diversity of *Prochlorococcus* in different environments (Kashtan et al. 2014, Malmstrom et al. 2013).

Because metagenomic techniques do not rely on a priori genetic information, they have been used to identify previously uncharacterized phylogenetic and functional phytoplankton biodiversity (Figures 4 and 5). The Global Ocean Sampling (GOS) survey represents the most extensive marine metagenomic study to date (Rusch et al. 2007), and several other studies have examined the phylogenetic diversity of *Prochlorococcus* and *Synechococcus* in samples from the GOS project (Huang et al. 2012, Rusch et al. 2010). These studies have revealed a phylogenetic clade of *Prochlorococcus* that is abundant in tropical high-nutrient, low-chlorophyll regions. A targeted metagenomic analysis of uncultured picoeukaryotic phytoplankton also enabled the assembly of a full genome sequence from a previously uncharacterized but abundant picoeukaryotic lineage (Cuvelier et al. 2010).

Metagenomic studies have also had a large impact on our understanding of the predicted functional diversity of phytoplankton, including previously unknown adaptations for photosynthesis, nutrient uptake and allocation, and other functions. For photosynthesis, a targeted metagenomic analysis identified the genome of the *ucynA* lineage of nitrogen fixers, which revealed that *ucynA* had lost photosystem II and thus is a photoheterotroph (Zehr et al. 2008). Furthermore, the lack of oxygen production explained the ability to fix nitrogen during the day. In a study of GOS samples, Larsson et al. (2014) identified novel configurations of the phycobilisome in *Synechococcus* that could provide an adaptation to low-salinity environments. Metagenomic studies have also revealed extensive variation in the presence or absence of nutrient acquisition genes. For example,

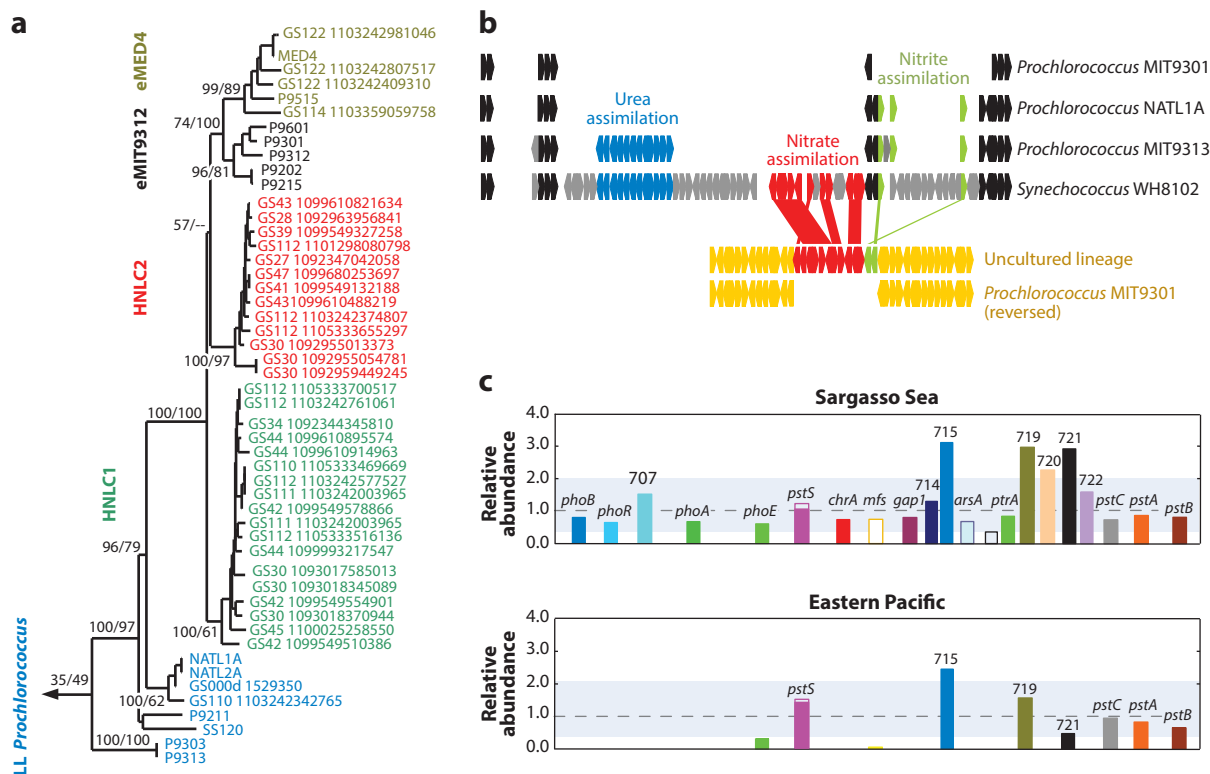


Figure 5

Examples of metagenomic analyses describing the diversity of *Prochlorococcus*. All analyses are based on metagenomic samples from the Global Ocean Sampling (GOS) survey (Rusch et al. 2007). (a) Identification of previously uncharacterized phylogenetic clades of *Prochlorococcus* found in high-nutrient, low-chlorophyll (HNLC) regions. The phylogenetic tree is based on sequences of the ribosomal protein rpsT and includes sequences from GOS as well as *Prochlorococcus* isolates. Data are from Rusch et al. (2010). (b) Linkage of *Prochlorococcus* to a cluster of nitrate assimilation genes. Analysis of metagenomic samples detected a previously unknown gene cluster in the genomes of uncultured *Prochlorococcus* ecotypes. Data are from Martiny et al. (2009b). (c) Phosphate acquisition gene frequencies in *Prochlorococcus* populations from areas of the Sargasso Sea with low phosphate concentrations and from areas of the eastern Pacific Ocean with high phosphate concentrations. The relative abundance of each gene was calculated as the ratio of the length-normalized occurrence of particular phosphate acquisition genes to the mean occurrence of all core *Prochlorococcus* genes (Kettler et al. 2007). Data are from Martiny et al. (2009a).

cells from regions with low nutrient concentrations contain a suite of genes for the uptake of nitrogen or phosphorus compounds, but those genes are generally absent in regions with elevated nutrient availability (Martiny et al. 2009a,b). This pattern has been observed in *Prochlorococcus* and *Synechococcus* as well as multiple heterotrophic bacteria and thus may be a general pattern for marine bacteria (Batnalle et al. 2014; Martiny et al. 2009a,b; Newton et al. 2010); by contrast, this pattern has not been detected in eukaryotic phytoplankton.

Metagenomic analyses have also been used to examine *Prochlorococcus* adaptations to low-iron conditions. Based on an assembly directly from a mixed population sample, the genome of the lineage appears to have a low frequency of genes with iron as a cofactor (Rusch et al. 2010). Malmstrom et al. (2013) analyzed single-cell genomes and found additional gene functions for utilizing organically bound iron; they therefore hypothesized that this lineage has adapted to low-iron conditions through a reduction in iron demand and an ability to take up alternative iron sources.

Functional diversity. Using cultured isolates in the laboratory, investigators can assay functional diversity through numerous approaches, from quantifying intrinsic growth rates to photosynthesis/respiration to studying behavior and more. Many of these assays can be modified for use in situ with mixed field communities, but there can be significant challenges because of low biomass/cell counts or complicating interactions from nonphotosynthetic organisms (e.g., respiration). Furthermore, many techniques are limited to bulk descriptions of the community, which, although useful for intercommunity comparisons, do not directly link biodiversity with function in the community.

Nevertheless, many approaches are available to describe functional diversity. For example, pigment labeling utilizes tracer-based techniques to quantify pigment synthesis and thus can be used to infer the growth rates of different taxonomic groups of phytoplankton (Goericke & Welschmeyer 1993). Similarly, HPLC or flow cytometry has been used in combination with grazing dilution experiments (or other incubation-type approaches) to quantify the growth and removal rates of specific phytoplankton taxa by leveraging the taxonomic resolution of the technique (Li 1994, Selph et al. 2011). Many of these taxonomically rooted techniques remain powerful tools for characterizing the functional diversity of mixed populations, but there have been additional advances that leverage recently available molecular approaches.

Gene expression: single loci. Functional diversity can be inferred from the abundance of transcribed genes by assuming that the level of transcription for a given gene is proportional to other forms of activity. For example, transcript levels of *nifH*, a gene involved in nitrogen fixation, are assumed to be proportional to nitrogen fixation and thus provide a molecular technique for assaying this function (Church et al. 2005). Indeed, for many microbial populations, functional activity is regulated at the transcriptional level, so there is a strong relationship between transcripts and functionality. Furthermore, noncoding RNA is emerging as a major component of functional regulation, thus providing significant insight into physiological diversity (Voigt et al. 2014). However, other functions—and, in particular, many functions associated with eukaryotic phytoplankton species—are not necessarily regulated at the transcriptional level, so care must be taken in interpretation. Furthermore, the relationships between molecular and physiological or biochemical activity (e.g., rRNA versus specific growth or diel variations in expression) can be complex and therefore must be used cautiously (Kerkhof & Kemp 1999, Lin et al. 2013, Zinser et al. 2009). Nevertheless, RNA-based assays can provide precise quantification of the responses of specific pathways, and there is often sufficient genetic resolution to differentiate the functional responses of different phytoplankton taxa. These properties have made this approach a powerful technique for assessing both specific pathways (e.g., *rbcL* for carbon fixation) and overall activity (e.g., rRNA) for both single species and mixed populations (Paul et al. 1999).

As with taxonomic diversity, two major approaches are often used when assessing gene expression (i.e., RNA). First, single genes (or gene fragments) can be quantified using qPCR of reverse-transcribed RNA. Here, known targets are quantified, providing a relatively precise level of transcription. As with qPCR or probe-based studies of taxonomic diversity, the major advantage of this approach is that the level of quantification is precise and there is a large dynamic range, but each group (species, clade, etc.) must be separately assessed. Sequencing of gene fragment libraries based on reverse-transcribed RNA is useful for mixed communities because it simultaneously assays the entire community, but these libraries are limited in their dynamic range and their precision of quantification. Because of these trade-offs, approaches based on reverse-transcription qPCR are typically used when there are relatively few known targets (e.g., a few clades) but many samples, whereas approaches based on sequencing are typically used when there are many targets or the number of targets is poorly constrained (Campbell et al. 2011). As sequencing costs decrease

and approaches become available for more precise quantification, sequence-based approaches are likely to become more common.

Gene expression: multiple loci. Similar to metagenomics, metatranscriptomics enables the diversity of phytoplankton functionality to be analyzed using RNA transcripts from the total community. This approach was first tested on a marine microbial community sampled near Hawaii (Frias-Lopez et al. 2008) and was later applied to many ocean regions and taxa (Allen et al. 2008, Marchetti et al. 2012, Poretsky et al. 2009, Stewart et al. 2010). The technique has multiple technical variants. Some of the challenges associated with this approach include (a) the amplification of RNA to obtain enough material for a sequencing library, (b) the removal of rRNA, and (c) the conversion of RNA into DNA. For eukaryotic RNA, the polyA tails of the transcripts are utilized to isolate the RNA, which is then linearly amplified and reverse transcribed into DNA. For prokaryotic or plastid RNA, the absence of a polyA tail presents a challenge. This problem can be solved by adding a polyA tail in vitro (Frias-Lopez et al. 2008, Poretsky et al. 2009) and then following the same steps as for eukaryotic phytoplankton. An alternative approach is to use random primers for reverse transcriptions and then perform MDA (Gilbert et al. 2008). Because rRNA typically constitutes the majority of transcripts, most studies utilize one of the (many) commercial kits available for removing (or at least reducing) rRNA (Luo et al. 2013). Finally, the abundance of cDNA can then be analyzed with either arrays or sequencing, as with metagenomics.

There are multiple caveats that are important to consider when interpreting metatranscriptomic data. As with any amplification technique, there is a risk of preferential amplification. This was evaluated for *Prochlorococcus*, where the profiles of amplified versus unamplified transcripts were correlated but had a large scatter ($r^2 = 0.85$ on log-transformed values) (Frias-Lopez et al. 2008). By contrast, the profiles of two amplified replicated samples can be similar, pointing to biased amplification. Other potential biases can be introduced in the rRNA removal step as well as through sequencing (He et al. 2007, Luo et al. 2013, Stewart et al. 2010). These observations suggest that metatranscriptomics is more accurate when used in comparisons, and that one should be cautious when interpreting the absolute transcript levels in a given sample.

Despite these caveats, metatranscriptomics can be a powerful technique for studying the diversity and functioning of phytoplankton in the ocean. For eukaryotes, one of the main advantages of metatranscriptomics over metagenomics is that it characterizes only exon regions. For example, Marchetti et al. (2009) used metatranscriptomics to identify genes in eukaryotic phytoplankton responding to iron stress. In prokaryotes, Shi et al. (2009) used the technique to identify novel transcripts (e.g., small RNAs), and Stewart et al. (2010) used it to identify differences in expression across environmental gradients. The approach can also identify temporal changes in gene expression in whole communities. As observed in many individual lineages, phytoplankton communities display a clear diel cycle in gene expression, whereas heterotrophic bacteria have very different gene expression patterns (Ottesen et al. 2011, 2014; Poretsky et al. 2009).

Proteomics. Metaproteomics is another technique that can be used to analyze phytoplankton community biodiversity. This technique is aimed primarily at assessing functional differences between communities or lineages because it represents the last step in the central dogma of biology (and therefore is presumably close to actual functionality). Traditionally, the proteins have been separated by two-dimensional page electrophoresis, with the profiles then compared or specific proteins extracted and further characterized. However, this technique has been largely replaced by mass spectrometry (Siggins et al. 2012). In short, current techniques require the isolation of proteins followed by a trypsin digest, which converts the proteins into polypeptides.

The peptides are then separated by gas and/or liquid chromatography in various combinations and then subjected to tandem mass spectrometry. This provides the mass of the peptide as well as the amino acids, which in turn are compared with a protein sequence database to determine the proteins in the sample. This database can be derived from sequenced genomes, metagenomes, or metatranscriptomes.

As with other techniques, there are multiple caveats. We know little about the biases introduced in the multiple preparation steps and the ability to fully identify the peptide sequences from the mass spectrometry analysis. Also, it is important to recognize that the protein identification is highly dependent on the sequence database. Thus, metaproteomics is most useful for well-characterized environments. However, as sequence information for marine communities rapidly accumulates, this may soon be an issue only for special marine environments. The technique is also limited by the number of total peptides identified, which restricts the depth and dynamic range of analysis. Furthermore, protein-based approaches are attractive because they are less removed from functionality, but proteins inherently have less information content compared with DNA and RNA, making them less useful for taxonomic/genomic studies.

Nevertheless, metaproteomics is becoming a powerful tool in marine biodiversity analyses. For example, one of the emerging findings is the high expression levels of transporters (Sowell et al. 2009, 2011), suggesting that microorganisms invest a considerable fraction of cellular resources into the uptake of nutrients and organic compounds. The technique has also been used to assess functional differences between regions (Morris et al. 2010) and between seasons (Williams et al. 2012).

OUTLOOK

As new techniques continue to develop and existing methods become more advanced, the tool kit available to characterize phytoplankton biodiversity will continue to expand. For example, DNA (and RNA) sequencing has now become relatively easy, such that the sample preparation and data analyses are a far larger cost (in both time and money) than the sequencing itself. Importantly, these advancements have also led to combinations of techniques that provide enhanced biodiversity resolution, such as the use of flow sorting and sequencing, dual RNA and DNA analyses, microscopy-based oligonucleotide primers of single cells, and many more.

Beyond the described “-omics” techniques, there have been major advances in the ability to directly quantify the diversity of biochemical capabilities among phytoplankton. Some of these techniques also rely on flow cytometry-based cell sorting and thus are restricted to lineages that can be identified based on size and/or fluorescence. An example is the ability to identify the nutrients that specific lineages can assimilate. For many years, it was thought that *Prochlorococcus* could not assimilate nitrate because all isolated cultures lacked this capability. However, using a combination of isotopically labeled substrates and cell sorting, Casey et al. (2007) showed that subpopulations of *Prochlorococcus* were indeed able to take up nitrate. A similar setup can enable investigators to estimate the quotas and ratios of cellular nutrients in populations of specific lineages; this approach revealed wide variation in the C:N:P ratios between and within *Prochlorococcus*, *Synechococcus*, and eukaryotic phytoplankton, which has significant implications for the link between the nutrient cycles and ratios in the ocean (Martiny et al. 2013). Other techniques are leveraging imaging with precision mass spectrometry to assay the uptake of labeled compounds by individual cells to directly quantify functionality (Woebken et al. 2012).

Investigators working on modern phytoplankton community characterization have a wide range of techniques at their disposal. These tools will continue to advance through ongoing improvements in both technology and the ability to interpret the often vast quantity of data generated.

DISCLOSURE STATEMENT

The authors are not aware of any affiliations, memberships, funding, or financial holdings that might be perceived as affecting the objectivity of this review.

ACKNOWLEDGMENTS

The authors thank all of those who provided comments on earlier versions of this review and Alyssa Kent for making **Figure 4**. Work in the authors' labs is supported by the US National Science Foundation (grants 0959630 and 103106 to Z.I.J. and 0928544, 1046297, and 1126749 to A.C.M.) and the US Department of Energy (grant DE-EE0003371 to Z.I.J.).

LITERATURE CITED

- Allen AE, LaRoche J, Maheswari U, Lommer M, Schauer N, et al. 2008. Whole-cell response of the pennate diatom *Phaeodactylum tricornutum* to iron starvation. *Proc. Natl. Acad. Sci. USA* 105:10438–43
- Amann R, Fuchs BM. 2008. Single-cell identification in microbial communities by improved fluorescence in situ hybridization techniques. *Nat. Rev. Microbiol.* 6:339–48
- Armbrust EV, Berges JA, Bowler C, Green BR, Martinez D, et al. 2004. The genome of the diatom *Thalassiosira pseudonana*: ecology, evolution, and metabolism. *Science* 306:79–86
- Batmalle CS, Chiang H-I, Zhang K, Lomas MW, Martiny AC. 2014. Development and bias assessment of a method for targeted metagenomic sequencing of marine cyanobacteria. *Appl. Environ. Microbiol.* 80:1116–25
- Benfield MC, Grosjean P, Culverhouse PF, Irigoien X, Sieracki ME, et al. 2007. RAPID: Research on Automated Plankton Identification. *Oceanography* 20(2):172–87
- Bertilsson S, Berglund O, Karl DM, Chisholm SW. 2003. Elemental composition of marine *Prochlorococcus* and *Synechococcus*: implications for the ecological stoichiometry of the sea. *Limnol. Oceanogr.* 48:1721–31
- Bitigore RR, Van Heukeleum L, Trees CC. 2005. Analysis of algal pigments by high-performance liquid chromatography. In *Algal Culturing Techniques*, ed. RA Andersen, pp. 327–45. Boston: Elsevier
- Blankenship RE. 2002. *Molecular Mechanisms of Photosynthesis*. Oxford, UK: Blackwell Sci.
- Bowler C, Allen AE, Badger JH, Grimwood J, Jabbari K, et al. 2008. The *Phaeodactylum* genome reveals the evolutionary history of diatom genomes. *Nature* 456:239–44
- Bragg LM, Stone G, Butler MK, Hugenholtz P, Tyson GW. 2013. Shining a light on dark sequencing: characterising errors in Ion Torrent PGM data. *PLoS Comput. Biol.* 9:e1003031
- Brown MV, Schwalbach MS, Hewson I, Fuhrman JA. 2005. Coupling 16S-ITS rDNA clone libraries and automated ribosomal intergenic spacer analysis to show marine microbial diversity: development and application to a time series. *Environ. Microbiol.* 7:1466–79
- Campbell BJ, Yu L, Heidelberg JF, Kirchman DL. 2011. Activity of abundant and rare bacteria in a coastal ocean. *Proc. Natl. Acad. Sci. USA* 108:12776–81
- Caporaso JG, Kuczynski J, Stombaugh J, Bittinger K, Bushman FD, et al. 2010. QIIME allows analysis of high-throughput community sequencing data. *Nat. Methods* 7:335–36
- Caporaso JG, Lauber CL, Walters WA, Berg-Lyons D, Lozupone CA, et al. 2011. Global patterns of 16S rRNA diversity at a depth of millions of sequences per sample. *Proc. Natl. Acad. Sci. USA* 108(Suppl. 1):4516–22
- Casey JR, Lomas MW, Mandecki J, Walker DE. 2007. *Prochlorococcus* contributes to new production in the Sargasso Sea deep chlorophyll maximum. *Geophys. Res. Lett.* 34:5
- Chandler DP, Fredrickson JK, Brockman FJ. 1997. Effect of PCR template concentration on the composition and distribution of total community 16S rDNA clone libraries. *Mol. Ecol.* 6:475–82
- Chitsaz H, Yee-Greenbaum JL, Tesler G, Lombardo M-J, Dupont CL, et al. 2011. Efficient de novo assembly of single-cell bacterial genomes from short-read data sets. *Nat. Biotechnol.* 29:915–21
- 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.* 71:5362–70

- Cordero OX, Polz MF. 2014. Explaining microbial genomic diversity in light of evolutionary ecology. *Nat. Rev. Microbiol.* 12:263–73
- Cuvelier ML, Allen AE, Monier A, McCrow JP, Messié M, et al. 2010. Targeted metagenomics and ecology of globally important uncultured eukaryotic phytoplankton. *Proc. Natl. Acad. Sci. USA* 107:14679–84
- DeLong EF, Preston CM, Mincer T, Rich V, Hallam SJ, et al. 2006. Community genomics among stratified microbial assemblages in the ocean's interior. *Science* 311:496–503
- Dubelaar GBJ, Gerritzen PL, Beeker AER, Jonker RR, Tangen K. 1999. Design and first results of CytoBuoy: a wireless flow cytometer for in situ analysis of marine and fresh waters. *Cytometry* 37:247–54
- Dyhrman ST, Jenkins BD, Rynearson TA, Saito MA, Mercier ML, et al. 2012. The transcriptome and proteome of the diatom *Thalassiosira pseudonana* reveal a diverse phosphorus stress response. *PLoS ONE* 7:e33768
- Escalas A, Bouvier T, Mouchet MA, Leprieur F, Bouvier C, et al. 2013. A unifying quantitative framework for exploring the multiple facets of microbial biodiversity across diverse scales. *Environ. Microbiol.* 15:2642–57
- Falkowski PG, Barber RT, Smetacek V. 1998. Biogeochemical controls and feedbacks on ocean primary production. *Science* 281:200–6
- Falkowski PG, Fenchel T, DeLong EF. 2008. The microbial engines that drive Earth's biogeochemical cycles. *Science* 320:1034–39
- Falkowski PG, Katz ME, Knoll AH, Quigg A, Raven JA, et al. 2004. The evolution of modern eukaryotic phytoplankton. *Science* 305:354–60
- Falkowski PG, Owens TG. 1980. Light-shade adaptation: two strategies in marine phytoplankton. *Plant Physiol.* 66:592–95
- Farrelly V, Rainey FA, Stackebrandt E. 1995. Effect of genome size and *rrn* gene copy number on PCR amplification of 16S rRNA genes from a mixture of bacterial species. *Appl. Environ. Microbiol.* 61:2798–801
- Field CB, Behrenfeld MJ, Randerson JT, Falkowski PG. 1998. Primary production of the biosphere: integrating terrestrial and oceanic components. *Science* 281:237–40
- Franks PJS. 1992. Sink or swim: accumulation of biomass at fronts. *Mar. Ecol. Prog. Ser.* 82:1–12
- Frias-Lopez J, Shi Y, Tyson GW, Coleman ML, Schuster SC, et al. 2008. Microbial community gene expression in ocean surface waters. *Proc. Natl. Acad. Sci. USA* 105:3805–10
- 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* 103:13104–9
- Gilbert JA, Field D, Huang Y, Edwards R, Li W, et al. 2008. Detection of large numbers of novel sequences in the metatranscriptomes of complex marine microbial communities. *PLoS ONE* 3:e3042
- Giovannoni SJ, Britschgi TB, Moyer CL, Field KG. 1990. Genetic diversity in Sargasso Sea bacterioplankton. *Nature* 345:60–63
- Goericke R, Welschmeyer NA. 1993. The chlorophyll-labeling method: measuring specific rates of chlorophyll *a* synthesis in cultures and in the open ocean. *Limnol. Oceanogr.* 38:80–95
- Haeckel E. 2005 (1862). *Art Forms from the Ocean: The Radiolarian Atlas of 1862*. Munich: Prestel
- Hall BG. 2011. *Phylogenetic Trees Made Easy: A How-To Manual*. Sunderland, MA: Sinauer
- He Z, Gentry TJ, Schadt CW, Wu L, Liebich J, et al. 2007. GeoChip: a comprehensive microarray for investigating biogeochemical, ecological and environmental processes. *ISME J.* 1:67–77
- Heldal M, Scanlan DJ, Norland S, Thingstad F, Mann NH. 2003. Elemental composition of single cells of various strains of marine *Prochlorococcus* and *Synechococcus* using X-ray microanalysis. *Limnol. Oceanogr.* 48:1732–43
- Hockin NL, Mock T, Mulholland F, Kopriva S, Malin G. 2011. The response of diatom central carbon metabolism to nitrogen starvation is different to that of green algae and higher plants. *Plant Physiol.* 158:299–312
- 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.* 6:285–97
- Hunt DE, Klepac-Ceraj V, Acinas SG, Gautier C, Bertilsson S, Polz MF. 2006. Evaluation of 23S rRNA PCR primers for use in phylogenetic studies of bacterial diversity. *Appl. Environ. Microbiol.* 72:2221–25
- Hunt DE, Lin Y, Church MJ, Karl DM, Izzo LK, et al. 2013. The relationship between abundance and specific activity of bacterioplankton in open ocean surface waters. *Appl. Environ. Microbiol.* 79:177–84

- Johnson ZI, Zinser ER, Coe A, McNulty NP, Woodward M, Chisholm SW. 2006. Niche partitioning among *Prochlorococcus* ecotypes along ocean-scale environmental gradients. *Science* 311:1737–40
- Karl DM. 2007. Microbial oceanography: paradigms, processes and promise. *Nat. Rev. Microbiol.* 5:759–69
- Kashtan N, Roggensack SE, Rodrigue S, Thompson JW, Biller SJ, et al. 2014. Single-cell genomics reveals hundreds of coexisting subpopulations in wild *Prochlorococcus*. *Science* 344:416–20
- Kerkhof L, Kemp P. 1999. Small ribosomal RNA content in marine Proteobacteria during non-steady-state growth. *FEMS Microbiol. Ecol.* 30:253–60
- Kettler GC, Martiny AC, Huang K, Zucker J, Coleman ML, et al. 2007. Patterns and implications of gene gain and loss in the evolution of *Prochlorococcus*. *PLoS Genet.* 3:e231
- Lang I, Hodac L, Friedl T, Feussner I. 2011. Fatty acid profiles and their distribution patterns in microalgae: a comprehensive analysis of more than 2000 strains from the SAG culture collection. *BMC Plant Biol.* 11:124
- Larsson J, Celepli N, Ininbergs K, Dupont CL, Yooseph S, et al. 2014. Picocyanobacteria containing a novel pigment gene cluster dominate the brackish water Baltic Sea. *ISME J.* 8:1892–903
- Li WKW. 1994. Primary production of prochlorophytes, cyanobacteria, and eukaryotic ultraphytoplankton: measurements from flow cytometric sorting. *Limnol. Oceanogr.* 39:169–75
- Lin Y, Gazsi K, Lance VP, Larkin A, Chandler J, et al. 2013. In situ activity of a dominant *Prochlorococcus* ecotype (eHL-II) from rRNA content and cell size. *Environ. Microbiol.* 15:2736–47
- Lomas MW, Bronk DA, van den Engh G. 2011. Use of flow cytometry to measure biogeochemical rates and processes in the ocean. *Annu. Rev. Mar. Sci.* 3:537–66
- Ludwig W, Strunk O, Westram R, Richter L, Meier H, et al. 2004. ARB: a software environment for sequence data. *Nucleic Acids Res.* 32:1363–71
- Luo C, Rodriguez RL, Konstantinidis KT. 2013. A user's guide to quantitative and comparative analysis of metagenomic datasets. *Methods Enzymol.* 531:525–47
- Mackey M, Mackey D, Higgins H, Wright S. 1996. CHEMTAX—a program for estimating class abundances from chemical markers: application to HPLC measurements of phytoplankton. *Mar. Ecol. Prog. Ser.* 144:265–83
- Maiden MCJ, Bygraves JA, Feil E, Morelli G, Russell JE, et al. 1998. Multilocus sequence typing: a portable approach to the identification of clones within populations of pathogenic microorganisms. *Proc. Natl. Acad. Sci. USA* 95:3140–45
- Malfatti F, Azam F. 2009. Atomic force microscopy reveals microscale networks and possible symbioses among pelagic marine bacteria. *Aquat. Microb. Ecol.* 58:1–14
- Malmstrom RR, Rodrigue S, Huang KH, Kelly L, Kern SE, et al. 2013. Ecology of uncultured *Prochlorococcus* clades revealed through single-cell genomics and biogeographic analysis. *ISME J.* 7:184–98
- Marchetti A, Parker MS, Moccia LP, Lin EO, Arrieta AL, et al. 2009. Ferritin is used for iron storage in bloom-forming marine pennate diatoms. *Nature* 457:467–70
- Marchetti A, Schruth DM, Durkin CA, Parker MS, Kodner RB, et al. 2012. Comparative metatranscriptomics identifies molecular bases for the physiological responses of phytoplankton to varying iron availability. *Proc. Natl. Acad. Sci. USA* 109:E317–25
- Mardis ER. 2013. Next-generation sequencing platforms. *Annu. Rev. Anal. Chem.* 6:287–303
- Martiny AC, Coleman ML, Chisholm SW. 2006. Phosphate acquisition genes in *Prochlorococcus* ecotypes: evidence for genome-wide adaptation. *Proc. Natl. Acad. Sci. USA* 103:12552–57
- Martiny AC, Huang Y, Li W. 2009a. Occurrence of phosphate acquisition genes in *Prochlorococcus* cells from different ocean regions. *Environ. Microbiol.* 11:1340–47
- Martiny AC, Kathuria S, Berube PM. 2009b. Widespread metabolic potential for nitrite and nitrate assimilation among *Prochlorococcus* ecotypes. *Proc. Natl. Acad. Sci. USA* 106:10787–92
- Martiny AC, Pham CTA, Primeau FW, Vrugt JA, Moore JK, et al. 2013. Strong latitudinal patterns in the elemental ratios of marine plankton and organic matter. *Nat. Geosci.* 6:279–83
- Meyer F, Paarmann D, D'Souza M, Olson R, Glass E, et al. 2008. The metagenomics RAST server—a public resource for the automatic phylogenetic and functional analysis of metagenomes. *BMC Bioinform.* 9:386
- Moeseneder MM, Arrieta JM, Muyzer G, Winter C, Herndl GJ. 1999. Optimization of terminal-restriction fragment length polymorphism analysis for complex marine bacterioplankton communities and comparison with denaturing gradient gel electrophoresis. *Appl. Environ. Microbiol.* 65:3518–25

- Moon-van der Staay SY, De Wachter R, Vaulot D. 2001. Oceanic 18S rDNA sequences from picoplankton reveal unsuspected eukaryotic diversity. *Nature* 409:607–10
- Morris RM, Nunn BL, Frazar C, Goodlett DR, Ting YS, Rocap G. 2010. Comparative metaproteomics reveals ocean-scale shifts in microbial nutrient utilization and energy transduction. *ISME J.* 4:673–85
- Newton RJ, Griffin LE, Bowles KM, Meile C, Gifford S, et al. 2010. Genome characteristics of a generalist marine bacterial lineage. *ISME J.* 4:784–98
- Olson RJ, Vaulot D, Chisholm SW. 1985. Marine phytoplankton distributions measured using shipboard flow cytometry. *Deep-Sea Res.* 32:1273–80
- Osanai T, Oikawa A, Shirai T, Kuwahara A, Iijima H, et al. 2014. Capillary electrophoresis–mass spectrometry reveals the distribution of carbon metabolites during nitrogen starvation in *Synechocystis* sp. PCC 6803. *Environ. Microbiol.* 16:512–24
- Ottesen EA, Marin R III, Preston CM, Young CR, Ryan JP, et al. 2011. Metatranscriptomic analysis of autonomously collected and preserved marine bacterioplankton. *ISME J.* 5:1881–95
- Ottesen EA, Young CR, Gifford SM, Eppley JM, Marin R, et al. 2014. Multispecies diel transcriptional oscillations in open ocean heterotrophic bacterial assemblages. *Science* 345:207–12
- Palenik B, Brahamsha B, Larimer FW, Land M, Hauser L, et al. 2003. The genome of a motile marine *Synechococcus*. *Nature* 424:1037–42
- Palenik B, Grimwood J, Aerts A, Rouze P, Salamov A, et al. 2007. The tiny eukaryote *Ostreococcus* provides genomic insights into the paradox of plankton speciation. *Proc. Natl. Acad. Sci. USA* 104:7705–10
- Palenik B, Ren Q, Tai V, Paulsen IT. 2009. Coastal *Synechococcus* metagenome reveals major roles for horizontal gene transfer and plasmids in population diversity. *Environ. Microbiol.* 11:349–59
- Parsons ML, Scholin CA, Miller PE, Doucette GJ, Powell CL, et al. 1999. *Pseudo-nitzschia* species (Bacillariophyceae) in Louisiana coastal waters: molecular probe field trials, genetic variability, and domoic acid analyses. *J. Phycol.* 35:1368–78
- Paul JH, Pichard SL, Kang JB, Watson GMF, Tabita FR. 1999. Evidence for a clade-specific temporal and spatial separation in ribulose biphosphate carboxylase gene expression in phytoplankton populations off Cape Hatteras and Bermuda. *Limnol. Oceanogr.* 44:12–23
- Pichard SL, Campbell L, Paul JH. 1997. Diversity of the ribulose biphosphate carboxylase/oxygenase form I gene (*rbcL*) in natural phytoplankton communities. *Appl. Environ. Microbiol.* 63:3600–6
- Polz MF, Cavanaugh CM. 1998. Bias in template-to-product ratios in multitemplate PCR. *Appl. Environ. Microbiol.* 64:3724–30
- Poretsky RS, Hewson I, Sun S, Allen AE, Zehr JP, Moran MA. 2009. Comparative day/night metatranscriptomic analysis of microbial communities in the North Pacific subtropical gyre. *Environ. Microbiol.* 11:1358–75
- Preheim SP, Perrotta AR, Martin-Platero AM, Gupta A, Alm EJ. 2013. Distribution-based clustering: using ecology to refine the operational taxonomic unit. *Appl. Environ. Microbiol.* 79:6593–603
- Read BA, Kegel J, Klute MJ, Kuo A, Lefebvre SC, et al. 2013. Pan genome of the phytoplankton *Emiliania* underpins its global distribution. *Nature* 499:209–13
- Rich VI, Konstantinidis K, DeLong EF. 2008. Design and testing of “genome-proxy” microarrays to profile marine microbial communities. *Environ. Microbiol.* 10:506–21
- Robidart JC, Preston CM, Paerl RW, Turk KA, Mosier AC, et al. 2012. Seasonal *Synechococcus* and *Thaumarochaeal* population dynamics examined with high resolution with remote in situ instrumentation. *ISME J.* 6:513–23
- Rocap G, Larimer FW, Lamerdin J, Malfatti S, Chain P, et al. 2003. Genome divergence in two *Prochlorococcus* ecotypes reflects oceanic niche differentiation. *Nature* 424:1042–47
- Rodrigue S, Malmstrom RR, Berlin AM, Birren BW, Henn MR, Chisholm SW. 2009. Whole genome amplification and de novo assembly of single bacterial cells. *PLoS ONE* 4:e6864
- Rusch DB, Halpern AL, Sutton G, Heidelberg KB, Williamson S, et al. 2007. The *Sorcerer II* Global Ocean Sampling expedition: Northwest Atlantic through Eastern Tropical Pacific. *PLoS Biol.* 5:e77
- Rusch DB, Martiny AC, Dupont CL, Halpern AL, Venter JC. 2010. Characterization of *Prochlorococcus* clades from iron-depleted oceanic regions. *Proc. Natl. Acad. Sci. USA* 107:16184–89
- Scanlan DJ, Hess WR, Partensky F, Newman J, Vaulot D. 1996. High degree of genetic variation in *Prochlorococcus* (Prochlorophyta) revealed by RFLP analysis. *Eur. J. Phycol.* 31:1–9

- Scanlan DJ, Ostrowski M, Mazard S, Dufresne A, Garczarek L, et al. 2009. Ecological genomics of marine picocyanobacteria. *Microbiol. Mol. Biol. Rev.* 73:249–99
- Schloss PD, Westcott SL, Ryabin T, Hall JR, Hartmann M, et al. 2009. Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Appl. Environ. Microbiol.* 75:7537–41
- Scholin C, Doucette G, Jensen S, Roman B, Pargett D, et al. 2009. Remote detection of marine microbes, small invertebrates, harmful algae, and biotoxins using the Environmental Sample Processor (ESP). *Oceanography* 22(2):158–67
- See JH, Campbell L, Richardson TL, Pinckney JL, Shen R, Guinasso NL. 2005. Combining new technologies for determination of phytoplankton community structure in the Northern Gulf of Mexico. *J. Phycol.* 41:305–10
- Selph KE, Landry MR, Taylor AG, Yang E-J, Measures CI, et al. 2011. Spatially-resolved taxon-specific phytoplankton production and grazing dynamics in relation to iron distributions in the equatorial Pacific between 110 and 140°W. *Deep-Sea Res. II* 58:358–77
- Shapiro HM. 2003. *Practical Flow Cytometry*. Hoboken, NJ: Wiley
- Shi Y, Tyson GW, DeLong EF. 2009. Metatranscriptomics reveals unique microbial small RNAs in the ocean's water column. *Nature* 459:266–69
- Shilova IN, Robidart JC, Tripp HJ, Turk-Kubo K, Wawrik B, et al. 2014. A microarray for assessing transcription from pelagic marine microbial taxa. *ISME J.* 8:1476–91
- Sieracki CK, Sieracki ME, Yentsch CS. 1998. An imaging-in-flow system for automated analysis of marine microplankton. *Mar. Ecol. Prog. Ser.* 168:285–96
- Siggins A, Gunnigle E, Abram F. 2012. Exploring mixed microbial community functioning: recent advances in metaproteomics. *FEMS Microbiol. Ecol.* 80:265–80
- Sogin ML, Morrison HG, Huber JA, Welch DM, Huse SM, et al. 2006. Microbial diversity in the deep sea and the underexplored “rare biosphere.” *Proc. Natl. Acad. Sci. USA* 103:12115–20
- Sowell SM, Abraham PE, Shah M, Verberkmoes NC, Smith DP, et al. 2011. Environmental proteomics of microbial plankton in a highly productive coastal upwelling system. *ISME J.* 5:856–65
- Sowell SM, Wilhelm LJ, Norbeck AD, Lipton MS, Nicora CD, et al. 2009. Transport functions dominate the SAR11 metaproteome at low-nutrient extremes in the Sargasso Sea. *ISME J.* 3:93–105
- Stewart FJ, Ottesen EA, DeLong EF. 2010. Development and quantitative analyses of a universal rRNA-subtraction protocol for microbial metatranscriptomics. *ISME J.* 4:896–907
- Thompson AW, Foster RA, Krupke A, Carter BJ, Musat N, et al. 2012. Unicellular cyanobacterium symbiotic with a single-celled eukaryotic alga. *Science* 337:1546–50
- Thompson JR, Marcelino LA, Polz MF. 2002. Heteroduplexes in mixed-template amplifications: formation, consequence and elimination by “reconditioning PCR.” *Nucleic Acids Res.* 30:2083–88
- Van Mooy BAS, Rocap G, Fredricks HF, Evans CT, Devol AH. 2006. Sulfolipids dramatically decrease phosphorus demand by picocyanobacteria in oligotrophic marine environments. *Proc. Natl. Acad. Sci. USA* 103:8607–12
- Venter JC, Remington K, Heidelberg JF, Halpern AL, Rusch D, et al. 2004. Environmental genome shotgun sequencing of the Sargasso Sea. *Science* 304:66–74
- Voigt K, Sharma CM, Mitschke J, Joke Lambrecht S, Voß B, et al. 2014. Comparative transcriptomics of two environmentally relevant cyanobacteria reveals unexpected transcriptome diversity. *ISME J.* 8:2056–68
- Westphal V, Rizzoli SO, Lauterbach MA, Kamin D, Jahn R, Hell SW. 2008. Video-rate far-field optical nanoscopy dissects synaptic vesicle movement. *Science* 320:246–49
- Williams TJ, Long E, Evans F, DeMaere MZ, Lauro FM, et al. 2012. A metaproteomic assessment of winter and summer bacterioplankton from Antarctic Peninsula coastal surface waters. *ISME J.* 6:1883–900
- Woebken D, Burow LC, Prufert-Bebout L, Bebout BM, Hoehler TM, et al. 2012. Identification of a novel cyanobacterial group as active diazotrophs in a coastal microbial mat using NanoSIMS analysis. *ISME J.* 6:1427–39
- Worden AZ, Lee J-H, Mock T, Rouze P, Simmons MP, et al. 2009. Green evolution and dynamic adaptations revealed by genomes of the marine picoeukaryotes *Micromonas*. *Science* 324:268–72
- Woyke T, Xie G, Copeland A, González JM, Han C, et al. 2009. Assembling the marine metagenome, one cell at a time. *PLoS ONE* 4:e5299

- Yentsch CM, Horan PK, Muirhead K, Dortch Q, Haugen E, et al. 1983. Flow cytometry and cell sorting: a technique for analysis and sorting of aquatic particles. *Limnol. Oceanogr.* 28:1275–80
- Zehr JP, Bench SR, Carter BJ, Hewson I, Niazi F, et al. 2008. Globally distributed uncultivated oceanic N₂-fixing cyanobacteria lack oxygenic photosystem II. *Science* 322:1110–12
- Zeidner G, Beja O. 2004. The use of DGGE analyses to explore eastern Mediterranean and Red Sea marine picophytoplankton assemblages. *Environ. Microbiol.* 6:528–34
- Zeidner G, Preston CM, Delong EF, Massana R, Post AF, et al. 2003. Molecular diversity among marine picophytoplankton as revealed by *psbA* analyses. *Environ. Microbiol.* 5:212–16
- Zhang K, Martiny AC, Reppas NB, Barry KW, Malek J, et al. 2006. Sequencing genomes from single cells by polymerase cloning. *Nat. Biotechnol.* 24:680–86
- Zinser ER, Coe A, Johnson ZI, Martiny AC, Fuller NJ, et al. 2006. *Prochlorococcus* ecotype abundances in the North Atlantic Ocean revealed by an improved quantitative PCR method. *Appl. Environ. Microbiol.* 72:723–32
- Zinser ER, Lindell D, Johnson ZI, Futschik M, Steglich C, et al. 2009. Choreography of the transcriptome, photophysiology, and cell cycle of a minimal photoautotroph, *Prochlorococcus*. *PLoS ONE* 4:e5135