# Genomes of Sea Microbes

BY MARY ANN MORAN AND E. VIRGINIA ARMBRUST

Genomics is the study of the genetic information encoded by all the nucleotides possessed by an organism. For oceanographers, genomics provides detailed information on the many genes that drive biogeochemical activities of ocean-dwelling microbes. Carbon fixation, nitrogen fixation, sulfur gas formation, $CO_2$ production, and many other critical processes are underlain by the collective action of genes inside individual microbial cells in the ocean environment. In essence, genomics provides access to those genes and serves as an important step toward understanding their role in the ocean environment.

The study of genes in the ocean had its start more than a decade ago when the new field of molecular ecology first allowed oceanographers to measure the diversity and distribution of selected protein-encoding genes. Polymerase chain reaction (PCR) amplification and sequencing of *nifH* genes led the way, providing information on the distribution of nitrogen-fixing organisms in marine environments and investigating the factors that limit their activity (Zehr and Capone, 1996). Similarly, studies of expression patterns of *rbcL*, the gene encoding a subunit of the major enzyme for carbon fixation, provided some of the first glimpses into cellular-level regulation of a major biogeochemical process in the ocean (Paul, 1996). These molecular oceanography studies were important predecessors to genomic oceanography studies, but they differ in two significant ways. First, the scale of genomics is grander, focusing on all the genes harbored by a marine organism simultaneously rather than just a few genes at a time. Second, genomics includes a large measure of discovery and does not require that the target genes be identified beforehand. Thus, it is not actually necessary to know what you are looking for in order to find genes that are novel or informative.

The difficulty of culturing every microbe in the sea and the dilemma of maintaining millions of microbial cultures simultaneously means that only a small fraction of marine microbial diversity can be studied as pure cultures. Selected model organisms have therefore begun to play an increasingly important role in oceanography. Because they can be grown in the laboratory and readily incorporated into experiments, ecologically relevant, culturable microbes provide a window into the ecology and physiology of their uncultured relatives. Put simply, they allow us to develop hypotheses about the uncultured multitude. The recent availability of genome sequences has greatly enhanced the value of model marine microbes, allowing manipulative experiments to be designed and interpreted much more specifically and insightfully. A few examples of sequenced marine microbes emerging as important model organisms in oceanography are: *Prochlorococcus marinus* (an abundant marine cyanobacterium; Rocap et al., 2003), *Pelagibacter ubique* (a member of the SAR11 clade; Giovannoni et al., 2005), *Silicibacter pomeroyi* (a Roseobacter; Moran et al., 2004), *Rhodopirellula*

For oceanographers, genomics provides detailed information on the many genes that drive biogeochemical activities of ocean-dwelling microbes.

MARY ANN MORAN *(mmoran@uga.edu) is Professor, Department of Marine Sciences, University of Georgia, Athens, GA, USA.* E. VIRGINIA ARMBRUST *is Professor, School of Oceanography, University of Washington, Seattle, WA, USA.*

*baltica* (a representative of the marine planctomycetes; Glöckner et al., 2003), *Thalassiosira pseudonana* (a marine diatom; Armbrust et al., 2004), and *Ostreococcus tauri* (a marine prasinophyte; Derelle et al., 2006).

"Pure-culture genomics" is distinct from but complementary to "metagenomics." In the latter (see Edwards and Dinsdale, this issue), organisms are sequenced without culturing and in the context of the other members of the community. In this chapter, we will focus on the former, addressing both the value of genomes of cultured marine microbes for tackling major questions in oceanography and the intrinsic synergism between pure-culture genomics and metagenomics.

## WHICH GENOMES SHOULD BE SEQUENCED?

DNA sequencing is becoming ever more rapid and inexpensive, putting marine microbial genomes within a few days' reach of a major sequencing center. Nonetheless, nontrivial postsequencing investments in genome analysis and annotation make organism selection a very important task.

Decisions about which marine prokaryotes to sequence began over a decade ago, with metabolic novelty serving as a major criterion. The hydrothermal-vent-dwelling, methane-generating *Methanocaldococcus jannaschii* was the first marine prokaryote sequenced (Bult et al., 1996), followed by the hyperthermophilic, sulfate-reducing *Archaeoglobus fulgidus* (Klenk et al., 1997) (Table 1). These genomes were used to gain insights into the novel ways marine microbes obtain energy

Table 1. Marine archaea and bacteria with completed genome sequences as of April 2007 (14 archaea, 47 bacteria).

| Organism | Genome[1] Size (MB) | ORFs[2] | Replicons[3] | rRNA Operons |
|---|---|---|---|---|
| ARCHAEA | | | | |
| *Aeropyrum pernix* K1 | 1.67 | 1841 | 1 | 1 |
| *Archaeoglobus fulgidus* DSM 4304 | 2.18 | 2420 | 1 | 1 |
| *Methanocaldococcus jannaschii* DSM 2661 | 1.74 | 1786 | 3 | 2 |
| *Methanococcus maripaludis* S2 | 1.66 | 1722 | 1 | 3 |
| *Methanopyrus kandleri* AV19 | 1.69 | 1687 | 1 | 1 |
| *Methanosarcina acetivorans* C2A | 5.75 | 4540 | 1 | 3 |
| *Nanoarchaeum equitans* Kin4-M | 0.49 | 536 | 1 | 1 |
| *Pyrobaculum aerophilum* IM2 | 2.22 | 2605 | 1 | 1 |
| *Pyrococcus abyssi* GE5 | 1.77 | 1898 | 2 | 1 |
| *Pyrococcus furiosus* DSM3638 | 1.91 | 2125 | 1 | 1 |
| *Pyrococcus horikoshii* OT3 | 1.74 | 1955 | 1 | 1 |
| BACTERIA | | | | |
| *Colwellia psychrerythraea* 34H | 5.37 | 4910 | 1 | 9 |
| *Desulfotalea psychrophila* LSv54 | 3.66 | 3234 | 3 | 7 |
| *Erythrobacter litoralis* HTCC2594 | 3.05 | 3011 | 1 | 1 |
| *Geobacillus kaustophilus* HTA426 | 3.59 | 3540 | 2 | 9 |
| *Hahella chejuensis* KCTC 2396 | 7.22 | 6778 | 1 | 5 |
| *Hyphomonas neptunium* ATCC 15444 | 3.71 | 3505 | 1 | 1 |
| *Idiomarina loihiensis* L2TR | 2.84 | 2628 | 1 | 4 |
| *Jannaschia sp.* CCS1 | 4.40 | 4283 | 2 | 1 |
| *Magnetococcus sp.* MC-1 | 4.70 | 3716 | 1 | 3 |
| *Maricaulis maris* MCS10 | 3.37 | 3063 | 1 | 2 |
| *Oceanobacillus iheyensis* HTE831 | 3.63 | 3500 | 1 | 7 |
| *Pelagibacter ubique* HTCC1062 | 1.30 | 1354 | 1 | 1 |
| *Photobacterium profundum* SS9 | 6.40 | 5491 | 3 | 15 |
| *Polaromonas sp.* JS666 | 5.90 | 5453 | 3 | 1 |
| *Prochlorococcus marinus* AS9601 | 1.70 | 1921 | 1 | 1 |
| *Prochlorococcus marinus* CCMP1375 (SS120) | 1.75 | 1882 | 1 | 1 |
| *Prochlorococcus marinus* MED4 | 1.66 | 1712 | 1 | 1 |
| *Prochlorococcus marinus* MIT 9301 | 1.60 | 1907 | 1 | 1 |

| Organism | Genome[1] Size (MB) | ORFs[2] | Replicons[3] | rRNA Operons |
|---|---|---|---|---|
| BACTERIA, continued | | | | |
| *Prochlorococcus marinus* MIT 9303 | 2.70 | 2997 | 1 | 2 |
| *Prochlorococcus marinus* MIT 9515 | 1.70 | 1906 | 1 | 1 |
| *Prochlorococcus marinus* MIT 9313 | 2.41 | 2265 | 1 | 1 |
| *Prochlorococcus marinus* NATL1A | 1.90 | 2193 | 1 | 1 |
| *Prochlorococcus marinus* NATL2A | 1.84 | 1892 | 1 | 1 |
| *Pseudoalteromonas atlantica* T6c | 5.19 | 4281 | 1 | 5 |
| *Pseudoalteromonas haloplanktis* TAC 125 | 3.85 | 3486 | 2 | 9 |
| *Psychromonas ingrahamii* 37 | 4.60 | 3545 | 1 | 10 |
| *Rhodopirellula baltica* SH 1 | 7.15 | 7325 | 1 | 1 |
| *Roseobacter denitrificans* | 4.30 | 4129 | 5 | 1 |
| *Saccharophagus degradans* 2-40 | 5.06 | 4008 | 1 | 2 |
| *Shewanella denitrificans* OS217 | 4.55 | 3754 | 1 | 8 |
| *Shewanella frigidimarina* NCIMB 400 | 4.85 | 4029 | 1 | 9 |
| *Shewanella sp.* MR-4 | 4.71 | 3924 | 1 | 9 |
| *Shewanella sp.* MR-7 | 4.80 | 4014 | 2 | 9 |
| *Shewanella sp.* PV-4 | 4.60 | 3859 | 1 | 8 |
| *Shewanella sp.* W3-18-1 | 4.70 | 4044 | 1 | 8 |
| *Silicibacter pomeroyi* DSS-3 | 4.60 | 4252 | 2 | 3 |
| *Silicibacter sp.* TM1040 | 4.15 | 3864 | 3 | 5 |
| *Sphingopyxis alaskensis* | 3.37 | 3195 | 2 | 1 |
| *Synechococcus sp.* (strain WH8102) | 2.43 | 2517 | 1 | 2 |
| *Synechococcus sp.* CC9311 | 2.61 | 2892 | 1 | 2 |
| *Synechococcus sp.* CC9605 | 2.51 | 2645 | 1 | 2 |
| *Synechococcus sp.* CC9902 | 2.23 | 2304 | 1 | 2 |
| *Thermotoga maritima* (strain MSB8) | 1.86 | 1858 | 1 | 1 |
| *Trichodesmium erythraeum* IMS101 | 7.75 | 4451 | 1 | 2 |
| *Vibrio fischeri* ES114 | 4.28 | 3802 | 3 | 12 |

[1]Draft genomes (those for which sequence gaps still remain) are not included.
[2]ORFs = open reading frames (regions likely to encode genes).
[3]Replicons = number of independent replicating DNA molecules, including chromosomes and plasmids.

and sense their environment. Many of the next wave of marine prokaryote genomes, however, were selected with ecological relevance rather than novelty as a primary criterion. Molecular taxonomic surveys of 16S rRNA genes revealed the identity of the major marine prokaryotic taxa (Giovannoni and Rappé, 2000; Suzuki and DeLong, 2002), and obtaining a genome sequence of representatives from as many of these groups as possible became a priority. This resulted in genome sequences from the Cyanobacteria, Roseobacter, SAR11, Flavobacteria, and Planctomycetes groups (Rocap et al., 2003; Moran et al., 2004; Giovannoni et al., 2005; Bauer et al., 2006; Glöckner et al., 2003) that are now being used to explore the genetic basis for ecological strategies and biogeochemically relevant activities of marine prokaryotes (Moran, in press).

The selection of marine eukaryotes for sequencing has generated much more heated discussion than selection of prokaryotes because of the larger scientific investment per genome. Relatively strict selection criteria had been developed to identify which medically relevant eukaryotes should be sequenced. For example, only organisms with a long history of being easily maintained and manipulated in numerous laboratories using molecular or genetic techniques were considered. But, the marine environment is characterized by incredibly diverse microbial communities and thus relatively few representative model marine microbes existed that could be easily manipulated in the laboratory. Instead, the strongest motivators were to identify the most ecologically important organisms and then put them in prior-

ity order based on genome size (smaller genomes of unicellular eukaryotes being preferred over larger genomes of multicellular eukaryotes) and ease of maintenance in the laboratory.

Because much of the early sequencing work on marine eukaroyotes was conducted through the Joint Genome Institute of the U.S. Department of Energy, the primary consideration for ecological importance was an organism's role in the global carbon cycle. Marine phytoplankton are responsible for about 45% of global carbon fixation on a yearly basis and, therefore, the common bloom-forming phytoplankton quickly came to the top of the list for consideration. Diatoms were considered first because they are estimated to be responsible for about 40% of marine primary productivity, followed by green algae, at least in part because of their relatively small genome sizes, and haptophytes (another group of algae) because of their role in precipitating an inorganic form of carbon (calcium carbonate). An obvious group of organisms missing from the list was and continues to be the dinoflagellates, which are perhaps best known as the species that synthesize very potent neurotoxins and bioluminescence. In general, dinoflagellates have extremely large genome sizes—many times larger than the human genome—and consequently most genomic work with dinoflagellates has relied upon sequencing only portions of genes that are expressed under different conditions, a technique that generates expressed sequence tags.

After weighing the options, the first marine eukaryote chosen for whole genome sequencing was the diatom *Thalassiosira pseudonana* (Armbrust et al., 2004). Since publication of this sequence, numerous other eukaryotic photosynthetic organisms have made their way into sequencing queues, and their whole genome sequences have either been completed or are well underway. The list includes:

- three additional diatoms—(1) *Phaeodactylum triconutum*, for which the most advanced tools for molecular biology and genetics are available, (2) *Pseudo-nitzschia multiseris*, which can produce the neurotoxin domoic acid and (3) *Fragilariopsis cylindrus*, which is restricted to polar environments.
- four green algae—(1) *Ostreococcus tauri* (Derelle et al., 2006), (2) *Ostreococcus lucimarinus*, which are bacterial-sized (about 1 micron in size) and are found in numerous coastal environments, and (3) two strains of *Micromonas pusilla* (see http://www.jgi.doe.gov), which is another important member of the eukaryotic picoplankton that bloom in both coastal and open-ocean water.
- the haptophyte *Emiliania huxleyi*, which plays a critical role in the production of calcium carbonate (Table 2).

In the past couple of years, sequencing technologies have advanced to the state where ecological and/or evolutionary relevance plays a much greater role than genome size in choosing which marine eukaryote to sequence next. For example, the genome of the choanoflagellate *Monosiga brevicollis* is currently being sequenced not only because it is among the closest unicellular relatives of animals, but also because of its role in transferring carbon to higher trophic levels in marine ecosystems (see http://www.jgi.doe.gov).

What will be the important criteria for choosing marine microbes for

Table 2. Marine eukaryotes with genome sequences as of April 2007.

| Organism | Group | Genome Size (Mb) | Gene Number | Status |
|---|---|---|---|---|
| *Ostreococcus tauri* | Prasinophyte | 12.6 | 8,166 | Finished |
| *Ostreococcus lucimarinus* | Prasinophyte | 13.2 | 7,651 | Finished |
| *Micromonas pusilla* (RCC299) | Prasinophyte | 15.0 | | Draft |
| *Micromonas pusilla* (CCMP 1545) | Prasinophyte | 15.0 | | Draft |
| *Bathycoccus sp.* (Ban7) | Prasinophyte | | | Underway |
| *Thalassiosira pseudonana* | Diatom | 34.0 | 11,242 | Finished |
| *Phaeodactylum tricornutum* | Diatom | 26.5 | 10,681 | Finished |
| *Pseudo-nitzschia multiseries* | Diatom | ~250.0 | | Underway |
| *Fragilariopsis cylindrus* | Diatom | ~35.0 | | Underway |
| *Aureococcus anophagefferens* | Pelagophyte | 32.0 | | Underway |
| *Emiliania huxleyi* | Coccolithophorid | ~220.0 | | Draft |
| *Guillardia theta* | Cryptomonad | | | Underway |
| *Monosiga brevicollis* | Choanoflagellate | 41.6 | 9,196 | Draft |

genome sequencing over the next several years? Organisms amenable to postgenomic studies (including development of a genetic system) are likely to be given high priority for the next wave of marine microbe sequencing, as molecular genetic and functional genomic methods facilitate improved access to gene function and regulation. Almost 50% of the genes encoded by many marine microbes do not display obvious similarity to genes from better-studied organisms. Therefore, determining the function of the proteins encoded by these "unknown" genes is a major priority. Among the marine microbial groups already represented by genome sequences, *Prochlorococcus*, *Synechococcus*, *Vibrio*, the Roseobacter group, diatoms, green algae, and haptophyes are good candidates for model organisms in support of marine functional genomics. For example, Roseobacter group member *S. pomeroyi* has been the focus of genomics-supported studies of novel voltage-dependent sodium ion channels (Koishi et al., 2004), sulfonate degradation (Cook et al., 2006), oxidation of lignin-related compounds (Buchan et al., 2004), and hydroxylamine oxidoreductases (Bergmann et al., 2005), among others. About 20 other Roseobacter genomes (completed or in progress) will soon be available for future comparative and functional genomic studies. The total number of sequenced eukaryotic microbes will continue to lag behind the number of sequenced prokaryotic microbes, but select groups of closely related species of marine phytoplankton are now being sequenced and will also be available for comparative studies.

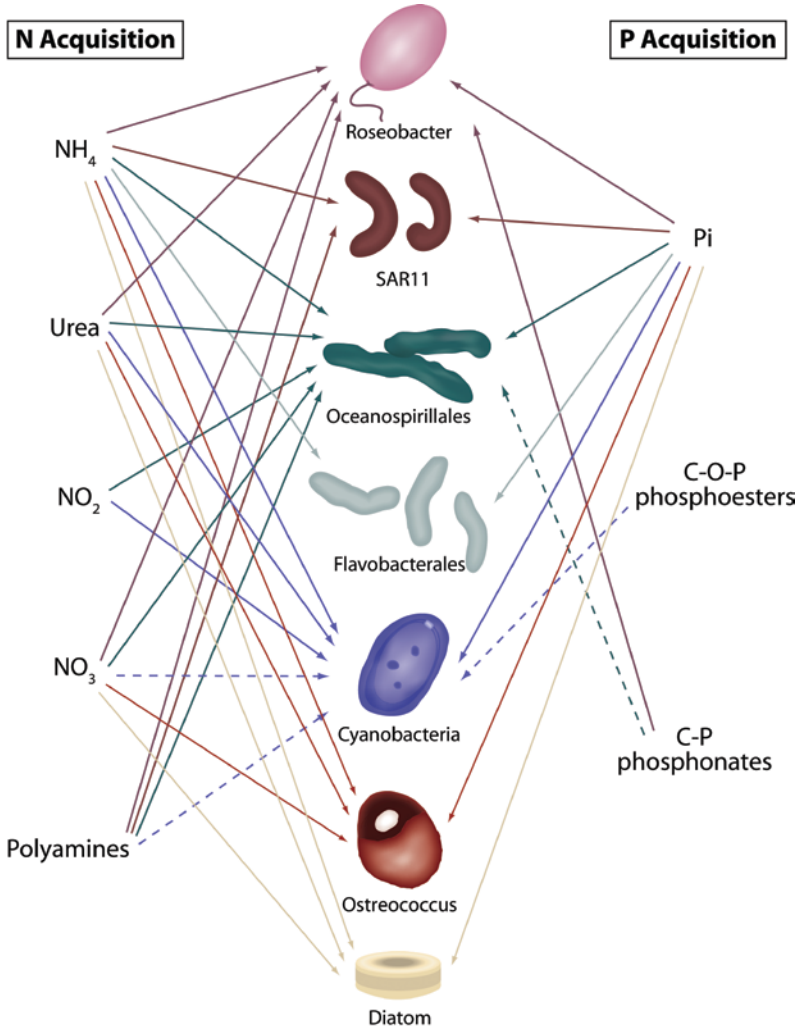


Figure 1. Hypothesized nitrogen and phosphorus utilization abilities for marine microbial taxa based on genes with sequence similarity to those whose role in N or P uptake and processing is known. Dashed line indicates utilization pattern may be variable within the group.

## DISCOVERING THE UNEXPECTED

The availability of whole genome sequences is particularly exciting because they serve as hypotheses-generating "machines." The piecing together of metabolic pathways or potential signaling pathways, for example, via computer-based analyses of genome sequences, can provide entirely new insights into how a particular group of organisms interacts with its environment (Figure 1). For example, analysis of the first diatom genome provided the first inklings that these organisms metabolized nitrogen in completely unexpected ways (Armbrust et al., 2004). These early analyses indicated that diatoms possessed a urea cycle, previously associated only with heterotrophic eukaryotes that use the pathway to excrete nitrogen waste resulting from consumption and breakdown of complex organic nitrogen compounds. The discovery of this pathway in a diatom, which requires inorganic nitrogen for growth, was entirely unanticipated despite years of intense focus

on nitrogen metabolism in marine phytoplankton. A second example comes from analysis of genomes of the green algae *Ostreococcus tauri* (Derelle et al., 2006) and *Ostreococcus lucimarinus* (Palenik et al., 2007). Computer-based analyses indicate that both organisms possess a chromosome with sequence features completely unlike those of other chromosomes, suggesting possible transfer of genes to both species from different organisms. This stunning result suggests that DNA may be more readily exchanged between eukaryotic organisms than previously suspected. A last example for marine eukaryotes has to do with the manner by which photosynthetic microorganisms convert carbon dioxide ($CO_2$) into organic carbon. In the ocean, most inorganic carbon is present as bicarbonate rather than $CO_2$. Because the enzyme that catalyzes the first step in the generation of organic carbon is specific for $CO_2$, most phytoplankton must somehow concentrate $CO_2$ intracellularly (Giordano et al., 2005). Careful analyses of multiple phytoplankton genomes now suggest that the C4 pathway, an alternative pathway for carbon fixation present in some land plants, may provide a mechanism for concentrating $CO_2$ (Derelle et al., 2006). This pathway is unexpected in marine phytoplankton, but may enhance their ability to fix carbon; thus, it has implications for understanding the global carbon cycle. This last example acts as a reminder that new discoveries about how marine microbes work can be found even in the most familiar places. Although these examples came from single eukaryotic organisms maintained in culture, they provide new avenues to explore in natural communities.

The value of genome sequences for revealing unexpected traits of microorganisms in the ocean has been demonstrated for prokaryotes as well. Rhodopsin genes in surface-water bacterioplankton (Béjà et al., 2000) and ammonia oxidation by marine archaea (Könneke et al., 2005) are two of the most dramatic examples of unanticipated discoveries of significant biogeochemical importance, although many others can be found. For example, discovering Type IV secretion systems in marine bacterial genomes was surprising (Moran et al., in press) because these systems are known to encode DNA export to eukaryotic cells (Dolowy et al., 2005), such as for the initiation of gall formation in the plant pathogen *Agrobacterium tumefaciens* (Christie et al., 2005). Because about half of the Roseobacter genomes contain Type IV secretion homologs, could these systems indicate widespread ability in this taxon to transfer DNA or protein to other plankton? A number of Roseobacter strains have been cultured in association with marine dinoflagellates (Alavi et al., 2001; Strompl et al., 2003), suggesting bacterial-dinoflagellate interactions as a possible role for these Type IV secretion systems (Worden et al., 2006).

> Pure-culture genomics has indeed begun to fundamentally change our understanding of who is doing what in the ocean, and how they are doing it.

## ENRICHING THE DETAILS

Along with the discovery of the unexpected, genome sequences of cultured marine microbes also provide new details about well-recognized processes. In a recent example, discovery of the genes mediating two critical steps in the sulfur cycle was made possible largely by pure-culture genomics. In these steps, an organic sulfur compound produced by marine phytoplankton (dimethylsulfoniopropionate, or DMSP) is converted by marine bacteria via one of two competing pathways (Kiene et al., 2000; Sievert et al., this issue). The first pathway leads to dimethylsulfide (DMS), a volatile sulfur compound that is readily transferred from the ocean to the atmosphere; the second leads to less-volatile compounds that are assimilated by bacteria (Figure 2). Which pathway dominates in marine surface waters has major implications for global temperature regulation because DMS exchanged across the ocean-atmosphere boundary affects cloud formation and global temperature, while non-DMS degradation products provide both sulfur and carbon to the marine microbial food web. Although the biochemical basis for the two competing pathways has heretofore been unknown, genome sequences of two cultured marine bacteria have recently led to identification of the

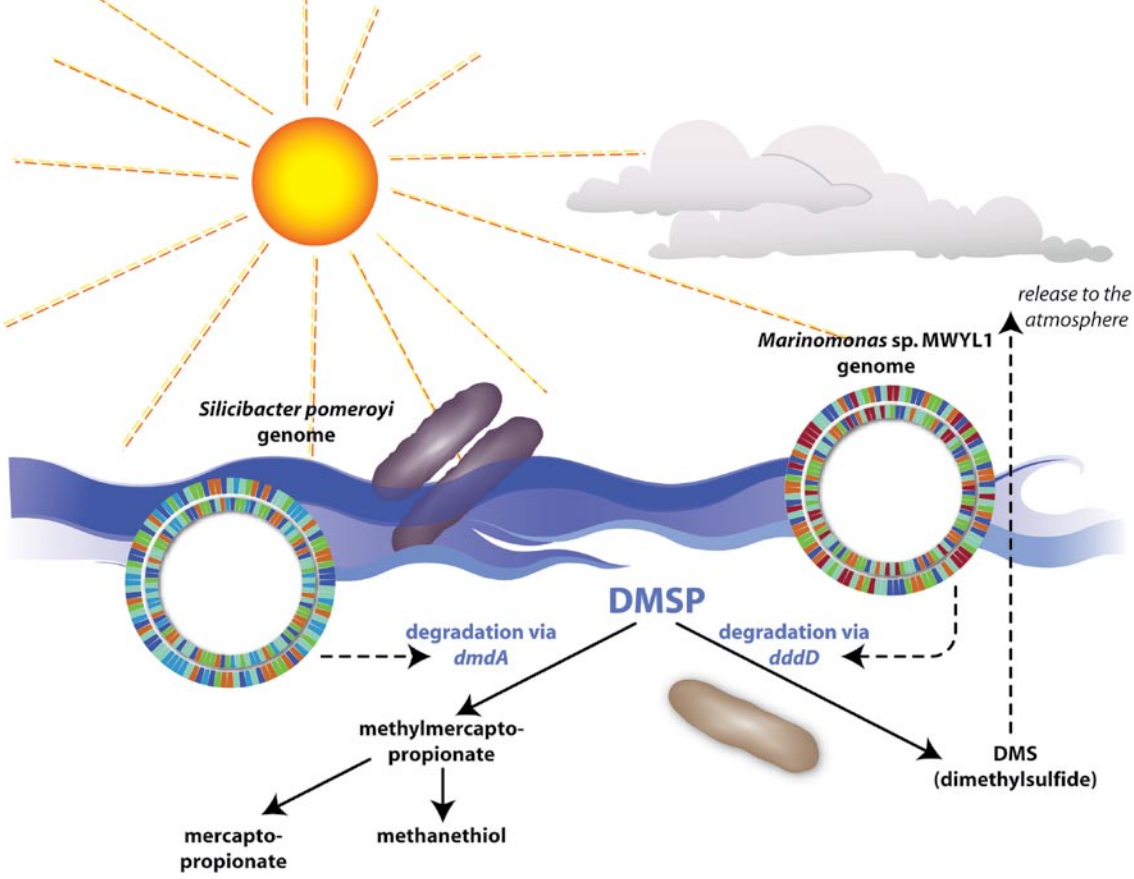


Figure 2. The availability of genome sequences of two cultured marine bacteria (*Silicibacter pomeroyi* and *Marinomonas* sp. MWYL1) led to the discovery of two key genes mediating dimethylsulfoniopropionate (DMSP) degradation. The genes encode the first enzyme in two competing pathways with vastly different ecological fates. The *dmdA* routes carbon and sulfur from DMSP to the marine microbial food web. The *dddD* degrades DMSP to a volatile sulfur compound that plays a critical role in the atmospheric sulfur pool.

DMSP demethylase gene (*dmdA*) that mediates DMSP degradation to non-DMS fates (Howard et al., 2006) and the DMSP cleavage gene (*dddD*) that mediates DMSP degradation to DMS (Todd et al., 2007) (Figure 2). With these two gene sequences in hand, understanding the regulation of this critical step in the marine sulfur cycle is on the horizon.

The silicon cycle is another important biogeochemical cycle, but the biology underlying silicon utilization remains poorly understood. Diatoms require silicon to produce elaborately patterned cell walls composed primarily of silicon, and they process about 7 billion metric tons of silicon on a yearly basis in doing so. Early insights into how diatoms use silicon to create their cell walls relied on painstaking biochemical analyses that by necessity were conducted with just one diatom, *Cylindrotheca fusiformis*. This work led to discovery of a novel class of lysine- and serine-rich phosphorylated proteins known as silaffins, which are involved in the precipitation of silica to create the diatom cell wall (Kröger et al., 1999). Analysis of the *T. pseudonana* genome not only identified additional silaffin genes (Poulsen and Kröger, 2004) but comparative analyses also highlighted important features of the encoded proteins. All the genes identified thus far that play a role in silicon utilization in diatoms are unique to these organisms, and identification of additional silicon-related genes will rely increasingly on the use of postgenomic techniques. These two examples, one from a prokaryote and another from a eukaryote, illustrate how pure-culture genomics opens windows into understanding the critical details of global-scale processes such as the sulfur and silicon cycles. Pure-culture genomics has indeed begun to fundamentally change our understanding of who is doing what in the ocean, and how they are doing it.

## LINKAGES TO METAGENOMICS

As the value of genomics to biological oceanography is becoming apparent, the next goal is to blend pure-culture genomics with metagenomics. The potential synergism of these two approaches is evident in the work of Coleman et al. (2006), who used the genome sequence of cultured *Prochlorococcus* strain MIT9312 to align sequence fragments from wild *Prochlorococcus* populations from the Sargasso Sea metagenome (Figure 3). Although most of the MIT9312 genome was well represented in the Sargasso Sea sequences, obvious gaps were found at "genomic islands." These may represent our first glimpse at the genetic basis for niche differentiation among *Prochlorococcus* species because the islands can provide a reservoir of interchangeable genes encoding ecologically important functions such as photoinhibition, nutrient uptake (amino acids, manganese/iron), nutri-

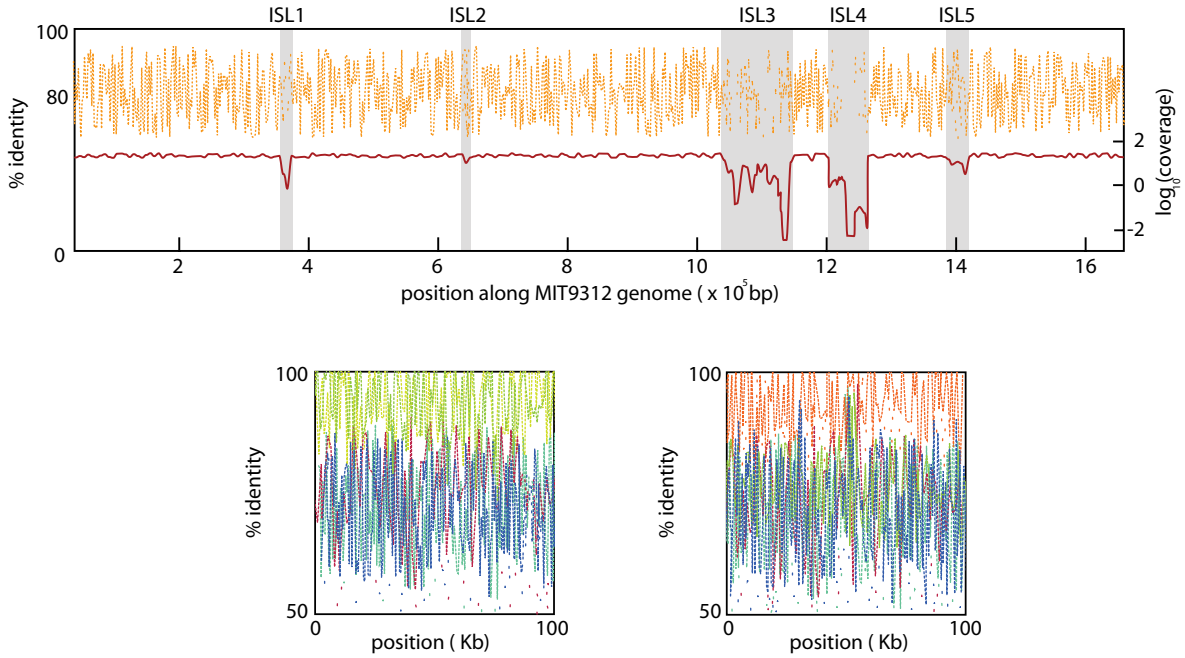


Figure 3. Synergism between pure-culture genomics and metagenomics is evident in genomic scaffolding analyses, in which metagenomic fragments are aligned against the genome sequence of a cultured marine microbe. Cultured organisms provide complete genome sequences and access to physiology, while metagenomic sequences show the abundance and distribution of genes and genome fragments from organisms that have not been cultured. Top: Metagenomic sequences from the Sargasso Sea were aligned to the genome sequence of cultured *Prochlorococcus* strain MIT9312 to show regions of shared genes typical of most *Prochlorococcus* as well as regions of variable genes (genomic islands indicated by shading) that may be important in defining the ecological niche of strains (*redrawn from Coleman et al., 2006*). Bottom: Metagenomic sequences from the Global Ocean Sampling expedition are aligned to the genome sequence of *Pelagibacter ubique* HTCC1062 to show population differences in coastal North American samples (yellow colors at left) versus Sargasso Sea samples (red colors at right) (*redrawn from Rusch et al., 2007*).

ent stress response, and phage resistance, among others (Coleman et al., 2006). A similar fragment recruitment approach was carried out with sequences from the larger Global Ocean Sampling expedition metagenome against several hundred available genomes of cultured prokaryotes. This research revealed closely related populations or "subtypes" of abundant marine prokaryotes, including *Pelagibacter*, *Prochlorococcus*, and *Synechococcus* (Rusch et al., 2007) (Figure 3). Some sequence differences among subtypes resulted in amino acid sequence changes in the encoded proteins, which might represent ecologically important differences in enzyme activity under varying environmental conditions.

The marriage between pure-culture genomics and metagenomics may ultimately be accomplished through the technology of single-cell genomics, in which individual cells are subjected to whole-genome sequencing. The genome of a single *Prochlorococcus* MIT9312 cell was recently obtained using "multiple displacement amplification," or MDA (Zhang et al., 2006). Although *Procholorococcus* MIT9312 is readily cultured in the laboratory, the leap to single-cell sequencing of uncultured cells plucked directly from the environment is obvious. Single-cell approaches have some of the advantages of pure-culture genomics, such as maintaining the packaging and regulation of ecological functions within individual cells, as well as some advantages of metagenomics, such as avoiding the significant biases imposed by a cultivation requirement. Functional genomic techniques appropriate for single cells are now clearly needed; these would allow access to physiology, biochemistry, and regulation of a single uncultured marine microbe.

## ACKNOWLEDGEMENTS

We thank C. English for preparing the figures, T. Mock and M. Parker for useful discussion, and Gordon and Betty Moore Foundation for grant support.

## REFERENCES


Alavi, M., T. Miller, K. Erlandson, R. Schneider, and R. Belas. 2001. Bacterial community associated with *Pfiesteria*-like dinoflagellate cultures. *Environmental Microbiology* 3:380–396.

Armbrust, E.V., J.A. Berges, C. Bowler, B.R. Green, D. Martinez, N.H. Putnam, S. Zhou, A.E. Allen, K.E. Apt, and others. 2004. The genome of the diatom *Thalassiosira pseudonana*: Ecology, evolution, and metabolism. *Science* 306:79–86.

Bauer, M., M. Kube, H. Teeling, M. Richter, T. Lombardot, E. Allers, C.A. Wurdemann, C. Quast, H. Kuhl, and others. 2006. Whole genome analysis of the marine Bacteroidetes '*Gramella forsetii*' reveals adaptations to degradation of polymeric organic matter. *Environmental Microbiology* 8:2,201–2,213.

Béjà, O., L. Aravind, E.V. Koonin, M.T. Suzuki, A. Hadd, L.H. Nguyen, S.B. Jovanovich, C.M. Gates, R.A. Feldman, J.L. Spudich, and others. 2000. Bacterial rhodopsin: Evidence for a new type of phototrophy in the sea. *Science* 289:1,902–1,906.

Bergmann, D.J., A.B. Hooper, and M.G. Klotz. 2005. Structure and sequence conservation of *hao* cluster genes of autotrophic ammonia-oxidizing bacteria: Evidence for their evolutionary history. *Applied and Environmental Microbiology* 71:5,371–5,382.

Buchan, A., E.L. Neidle, and M.A. Moran. 2004. Diverse organization of genes of the ß-ketoadipate pathway in members of the marine Roseobacter lineage. *Applied and Environmental Microbiology* 70:1,658–1,668.

Bult, C.J., O. White, G.J. Olsen, L. Zhou, R.D. Fleischmann, G.G. Sutton, J.A. Blake, L.M. FitzGerald, R.A. Clayton, J.D. Gocayne, and others. 1996. Complete genome sequence of the methanogenic archaeon, *Methanococcus jannaschii*. *Science* 273:1,058–1,073.

Christie, P.J., K. Atmakuri, V. Krishnamoorthy, S. Jakubowski, and E. Cascales. 2005. Biogenesis, architecture, and function of bacterial Type IV secretion systems. *Annual Review of Microbiology* 59:451–485.

Coleman, M.L., M.B. Sullivan, A.C. Martiny, C. Steglich, K. Barry, E.F. DeLong, and S.W. Chisholm. 2006. Genomic islands and the ecology and evolution of Prochlorococcus. *Science* 311:1,768–1,770.

Cook, A.M., K. Denger, and T.H.M. Smits. 2006. Dissimilation of C3-sulfonates. *Archives of Microbiology* 185:83–90.

Derelle, E., C. Ferraz, S. Rombauts, P. Rouze, A.Z. Worden, S. Robbens, F. Partensky, S. Degroeve, S. Echeynié, R. Cooke, and others. 2006. Genome analysis of the smallest free-living eukaryote *Ostreococcus tauri* unveils many unique features. *Proceedings of the National Academy of Sciences of the United States of America* 103:11,647–11,652.

Dolowy, P., J. Mondzelewski, R. Zawadzka, J. Baj, and D. Bartosik. 2005. Cloning and characterization of a region responsible for the maintenance of megaplasmid pTAV3 of *Paracoccus versutus* UW1. *Plasmid* 53:239–250.

Giordano, M., J. Beardall, and J.-A. Raven. 2005. $CO_2$ concentrating mechanisms in algae: Mechanisms, environmental modulation, and evolution. *Annual Review of Plant Biology* 56:99–131.

Giovannoni, S.J., and M.S. Rappé. 2000. The uncultured microbial majority. Pp. 47–84 in *Microbial Ecology of the Oceans.* D.L. Kirchman, ed., Wiley-Liss, New York.

Giovannoni, S.J., H.J. Tripp, S. Givan, M. Podar, K.L. Vergin, D. Baptista, L. Bibbs, J. Eads, T.H. Richardson, M. Noordewier, and others. 2005. Genome streamlining in a cosmopolitan oceanic bacterium. *Science* 309:1,242–1,245.

Glöckner, F.O., M. Kube, M. Bauer, H. Teeling, T. Lombardot, W. Ludwig, D. Gade, A. Beck, K. Borzym, K. Heitmann, and others. 2003. Complete genome sequence of the marine planctomycete *Pirellula* sp. strain 1. *Proceedings of the National Academy of Sciences of the United States of America* 100:8,298–8,303.

Howard, E.C., J.R. Henriksen, A. Buchan, C.R. Reisch, H. Bürgmann, R. Welch, W. Ye, J.M. González, K. Mace, S.B. Joye, and others. 2006. Bacterial taxa that limit sulfur flux from the ocean. *Science* 314:649–652.

Kiene, R.P., L.J. Linn, and J.A. Bruton. 2000. New and important roles for DMSP in marine microbial communities. *Journal of Sea Research* 43:209–224.

Klenk, H.P., R.A. Clayton, J.-F. Tomb, O. White, K.E. Nelson, K.A. Ketchum, R.J. Dodson, M. Gwinn, E.K. Hickey, J.D. Peterson, and others. 1997. The complete genome sequence of the hyperthermophilic, sulphate-reducing archaeon *Archaeoglobus fulgidus*. *Nature* 390:364–370.

Koishi, R., H. Xu, D. Ren, B. Navarro, B.W. Spiller, Q. Shi, and D.E. Clapham. 2004. A superfamily of voltage-gated sodium channels in bacteria. *Journal of Biological Chemistry* 279:9,532–9,538.

Könneke, M., A.E. Bernhard, J.R. de La Torre, C.B. Walker, J.B. Waterbury, and D.A. Stahl. 2005. Isolation of an autotrophic ammonia-oxidizing marine archaeon. *Nature* 437:543–546.

Kröger, N., R. Deutzmann, and M. Sumper. 1999. Polycationic peptides from diatom biosilica that direct silica nanosphere formation. *Science* 286:1,129–1,132.

Moran, M.A. In press. Marine prokaryotic genomics and metagenomics. In *Microbial Ecology of the Oceans*. D.L. Kirchman, ed., Wiley, New York.

Moran, M.A., A. Buchan, J.M. González, J.F. Heidelberg, W.B. Whitman, R.P. Kiene, J.R. Henriksen, G.M. King, R. Belas, C. Fuqua, and others. 2004. Genome sequence of *Silicibacter pomeroyi* reveals adaptations to the marine environment. *Nature* 432:910–913.

Moran, M.A., R. Belas, M.A. Schell, J.M. González, F. Sun, S. Sun, B.J. Binder, J. Edmonds, W. Ye, B. Orcutt, E. Howard, and others. In press. Ecological genomics of marine roseobacters. *Applied and Environmental Microbiology*.

Palenik, B., J. Grimwood, A. Aerts, P. Rouzé, A. Salamov, N. Putnam, C. Dupont, R. Jorgensen, E. Derelle, S. Rombauts, and others. 2007. The tiny eukaryote *Ostreococcus* provides genomic insights into the paradox of plankton speciation. *Proceedings of the National Academy of Sciences of the United States of America* 104:7,705–7,710.

Paul, J.H. 1996. Carbon cycling: Molecular regulation of photosynthetic carbon fixation. *Microbial Ecology* 32:231–245.

Poulsen, N., and N. Kröger. 2004. Silica morphogenesis by alternative processing of silaffins in the diatom *Thalassiosira pseudonana. Journal of Biological Chemistry* 279:42,993–42,999.

Rocap, G., F.W. Larimer, J. Lamerdin, S. Malfatti, P. Chain, N.A. Ahlgren, A. Arellano, M. Coleman, L. Hauser, W.R. Hess, and others. 2003. Genome divergence in two *Prochlorococcus* ecotypes reflects oceanic niche differentiation. *Nature* 424:1,042–1,047.

Rusch, D.B., A.L. Halpern, G. Sutton, K.B. Heidelberg, S. Williamson, S. Yooseph, D. Wu, J.A. Eisen, J.M. Hoffman, K. Remington, and others. 2007. The *Sorcerer II* Global Ocean Sampling expedition: Northwest Atlantic through Eastern Tropical Pacific. *PLoS Biology* 5:e77 doi:10.1371/journal.pbio.0050077.

Strömpl, C., G.L. Hold, H. Lünsdorf, J. Graham, S. Gallacher, W.-R. Abraham, E.R.B. Moore, and K.N. Timmis. 2003. *Oceanicaulis alexandrii* gen. nov., sp. nov., a novel stalked bacterium isolated from a culture of the dinoflagellate *Alexandrium tamarense* (Lebour) Balech. *International Journal of Systematic and Evolutionary Microbiology* 53:1,901–1,906.

Suzuki, M.T., and E.F. DeLong. 2002. Marine prokaryote diversity. Pp. 209–234 in *Biodiversity of Microbial Life*. J.T. Staley and A.-L. Reysenbach, eds, Wiley-Liss, New York.

Todd, J., R. Rodgers, Y.G. Li, M. Wexler, P.L. Bond, L. Sun, A.R.J. Curson, G. Malin, M. Steinke, and A.W.B. Johnston. 2007. Structural and regulatory genes required to make the gas dimethyl sulfide in bacteria. *Science* 315:666–669.

Worden, A.Z., M.L. Cuvelier, and D.H. Bartlett. 2006. In-depth analyses of marine microbial community genomics. *Trends in Microbiology* 14:331–336.

Zehr, J.P., and D.G. Capone. 1996. Problems and promises of assaying the genetic potential for nitrogen fixation in the marine environment. *Microbial Ecology* 32:263–281.

Zhang, K., A.C. Martiny, N.B. Reppas, K.W. Barry, J. Malek, S.W. Chisholm, and G.M. Church. 2006. Sequencing genomes from single cells by polymerase cloning. *Nature Biotechnology* 24:680–686.